SIMULATION STUDIES OF THE ORGANIC SOLID STATE

Antonio Torrisi

A thesis submitted in partial fulfilment of the requirements of the University of London for the degree of Doctor of Philosophy, July 2007.

Davy Faraday Research Laboratory (DFRL)
The Royal Institution of Great Britain
Kathleen Lonsdale Building
Gower Street
London WC1E 6BT

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

"Ti proteggerò dalle paure delle ipocondrie,
dai turbamenti che da oggi incontrerai per la tua via.
Dalle ingiustizie e dagli inganni del tuo tempo,
dai fallimenti che per tua natura normalmente attirerai.
Ti solleverò dai dolori e dai tuoi sbalzi d'umore,
dalle ossessioni delle tue manie.
Supererò le correnti gravitazionali,
lo spazio e la luce
per non farti invecchiare.
E guarirai da tutte le malattie,
perché sei un essere speciale,
ed io, avrò cura di te.

Vagavo per i campi del Tennessee
(come vi ero arrivato, chissà).
Non hai fiori bianchi per me?
Più veloci di aquile i miei sogni
attraversano il mare.

Ti porterò soprattutto il silenzio e la pazienza.
Percorreremo assieme le vie che portano all'essenza.
I profumi d'amore inebrieranno i nostri corpi,
la bonaccia d'agosto non calmerà i nostri sensi.
Tesserò i tuoi capelli come trame di un canto.
Conosco le leggi del mondo, e te ne farò dono.
Supererò le correnti gravitazionali,
lo spazio e la luce per non farti invecchiare.
Ti salverò da ogni malinconia,
perché sei un essere speciale ed io avrò cura di te...
io sì, che avrò cura di te."

(*"La Cura"* by Franco Battiato)

To my father, my mother, and all my friends, who have been near me in my life...

Abstract

Antonio Torrisi

This thesis concerns the investigation of the organic solid state from a computational point of view, by using classical simulation methods to predict and study the structure and physical properties of simple organic properties. The initial two chapters of the thesis report an exhaustive introduction about the background, which the computational work is based on: polymorphism in the organic solid state (chapter 1), phase transition in the solid state (chapter 2). This is followed by a description of the theoretical methods used in the work (chapter 3). The results reported in the thesis are mainly divided in two parts: the first one is an application of one of the current computational method for crystal structure prediction (CSP) of organic compounds to three simple molecules: imidazole (chapter 4), an intrinsically rigid molecule with high polar groups, azetidine (chapter 5), a simple molecule with a moderate internal flexibility in the molecular ring and with the presence of a polar group and cyclopentane (chapter 6), a simple molecule with internal flexibility in the molecular ring and lack of polar groups. The work on azetidine was carried out as part of the international blind test in CSP. The second part of the results concerns the use of Molecular Dynamics (MD) technique in order to studies structural changes of crystal structures with temperature. The main part of the work concerns the study of the orientationally *order-disorder* phase transition in cyclopentane solid phase (chapter 7). Finally a MD study of a limited number of imidazole structures was performed in order to estimate the temperature effect on the structures generated by the CSP procedure (chapter 4). The thesis demonstrates the complex nature and behaviour characterising the organic solid state, even for very simple molecules.

Acknowledgements

In performing and completing this long study, I did not, and I could not even, do all alone, but with the precious help of many people around me. Life is not to be alone. Hence I would like to thank first of all my supervisors Prof. Richard Catlow and Prof. Sally Price, for their continuous suggestions, support and advices and their help in my PhD work and during the writing of this thesis.

I would also like to thank Dr. Maurice Leslie and Dr. William Smith at Daresbury Laboratory, for their big and important help and advices. A huge thank also to Dr. David Coombes, for his enormous support and help during all my work and for the uncountable conversations we had about everything, from material science to Queen. A very big thank to my dear friend Dr. Said Hamad Gomez and to Dr. Panos Karamertzanis for their precious advices and helps during my work and big thank to the people in the Sally's group, Dr. Charlie Leech, Dr. Tom Lewis, Dr. Harriot Nowell and those in the RI group, Dr. Devis Di Tommaso, Dr. Alexei Sokol, Gareth Tribello, Malek Deifallah, for suggestions, discussions and corrections in writing my thesis.

I would like to thank my Spanish friends, Alicia, Rocio & Alessia, Hector, Miguel G., Luis, Raquel, the great Cuban Ricardo Grau Crespo, Katherine for all the support received by them during these years.

A huge thanks to my girlfriend Jehanne Guillet, for the big support she gave me and her continuous presence in this year.

I would like to thanks dear Giuseppe Mallia for his help and support in the most difficult moments and Furio Cora' e Maria Alfredsson, for their encouragements.

Finally I would like to thank my friends in Italia, Michele, Albina, Danilo, Enrico T., Fabio, Marco S., Claudia, Fabri, Emanuele, Pino, Andrea, Lorenzo, Federica, my dear Romanian friend Antoaneta and my Brazilian friend Rafael, Gidi and Kamilla.

Ultimately I would like to thank immensely my father and my mother, who have been very close to me, although the distance, and never stopped to encourage and support me.

Table of Contents

List of Tables and Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 1

POLYMORPHISM & CRYSTAL STRUCTURE PREDICTION IN THE ORGANIC SOLID STATE

1.1 Introduction

1.1.1 Polymorphism and Pseudo-polymorphism.

Polymorphism is a purely solid state effect, for which there are different definitions. McCrone gave the following in 1965[1]:"*A polymorph is a solid crystalline phase of a given compound, resulting from the possibility of at least two different arrangements of the molecules of that compound in the solid state.*" We can consider polymorphs as different solids or different crystals of the same compound, which are obtained from different arrangements of the same molecule in the space. An example of a polymorphic compound is carbon, which exists in three different polymorphic solid forms, which are diamond, graphite and fullerene. Polymorphism is well known in compounds of inorganic materials, e.g. zinc sulphide (ZnS) or calcium carbonate ($CaCO_3$), but is also extensively present in the organic solid state as well, where one of the most illustrative examples is given by benzene, which has seven possible polymorphs[2]. Many other important organic compounds, i.e. soaps[3], waxes[4], pigments[5], liquid crystals[6] and drugs[7,8] have many well characterised polymorphs.

Crystals are characterised by the order and regularity of the structures in which the molecule is arranged. Nevertheless there are also amorphous forms, which are considered polymorphs. Amorphous phases are important, because of their metastability, which makes them more soluble than the crystalline polymorphs of a certain compound[9]. Moreover the amorphous state itself can be characterised by the presence of polymorphism (which in this case takes the name of *polyamorphism*), as shown by studies of polymorphism of amorphous solids at high pressure[10].

In the solid state there is a distinction between polymorphism and polytypism. The latter is a 1D-case of polymorphism, consisting of different stackings of the same layers of a compound. It is widespread in silicon carbide SiC and also in the organic solid state[11], both as ordered[12] and as disordered stacking[13]. Polymorphism is present in many different organic compounds, but also in proteins[14,15] (referring both to minor molecular

conformational changes and to different packing), thin films[16], liquid crystals[17] and polymers[18].

The proper definition of polymorphism among organic solid compounds and the criteria to use in establishing the real polymorphic nature of a specific phase with respect to another for a specific compound is for years the subject of debate and discussions[19] and they are still now not completely clear. It is difficult to establish whether a certain compound is polymorphic or not. The issue about how many compounds in the organic solid state are polymorphic is still open[20] and in a paper in 1969 McCrone[21] made a provocative statement about the occurrence of polymorphism: "*Every compound has different polymorphic forms and the number of forms known for a given compound is proportional to the time and energy spent in research on that compound*". Statistical analysis of the occurrence of polymorphism is difficult because many instances of polymorphism are not properly reported. The *Cambridge Structure Database* (CSD)[22], is a database of crystal structures of small organic molecules and metal-organic compounds, reporting the structural information[22]. A survey of the CSD on the keywords of "polymorphs", "forms", "phase", and "modification" indicates that about 4% of the compounds contained in the database are polymorphic[20]. This outcome depends on the fact that it includes only compounds for which the crystal structures have been reported. On the other hand, another recent work on over 150 compounds showed that more than 87% of them has more than one form[23]. The realistic situation is probably in between these two extreme examples as expressed by Bernstein in one recent paper[20]: "*The existence of multiple crystal forms is not surprising when found, but cannot be foreseen or expected for any compound*".

There are different kinds of polymorphism as well as different forms of certain compounds, for which the definition of polymorphism rather than the term of *pseudo-polymorphism*, have been strongly debated by different experts (McCrone[1], Dunitz[13], Threfall[11]). We can divide the phenomenon of polymorphism into the following different classes:

a*) Packing Polymorphism*

This kind of polymorphism represents situations in which rigid molecules can assembled in different three-dimensional structures and arrangements, through different intermolecular interactions[24]. An example is provided by *p*-nitrophenol, for which the two polymorphs contain both four crystallographic independent molecules in the unit cell, linked by dimers through hydrogen bonding[25]. Another typical example of packing

polymorph exists in sulfathiazole[26], with the differences in packing type yielding the different forms. This is also observed in the solid state of terephthalic acid, which is characterised by the presence of three different polymorphs[27], with two different forms (conventionally called form I and form II), both with a triclinic unit cell and the form III with a monoclinic unit cell. In the two triclinic forms the molecules are linked by hydrogen-bonded carboxyl dimers, into infinite chains forming two-dimensional layers. The main difference between these forms concerns the arrangement of the chains.

A very common case of packing polymorphism depends on different possible hydrogen bond motifs in the crystal structure. Many compounds of molecules containing polar groups, such as O-H, N-H, C=O, crystallise by forming particular networks connected through hydrogen bond interactions, such as N-H···O, N-H···N, H-O···H, C=O···H, C=O···H-N. For a certain compound we can observe different polymorphs, characterised by different arrangements, or "motifs", of the possible inter and intra-molecular hydrogen bonds between different intermolecular groups. A typical example of this polymorphism is found in oxalic acid[28], which has two polymorphs: the alpha-form having a crinkled sheet structure, with $O^{(2)}$-H···$O^{(1)}$ hydrogen bonds forming a chain, and the beta-form characterised by chains of carboxylic acid dimers of a particular R^2_2 (8) graph set[29]. Many of these hydrogen bond patterns were characterised, described and classified[29]. Typical hydrogen bond patterns are dimers in mono and di-carboxylic acids, and catemers (see figure 1.1), which often are in competition with each other in determining the most stable polymorphic form for a compound with many molecular polar groups, as investigated in a computational study[30] of these hydrogen bond patterns.

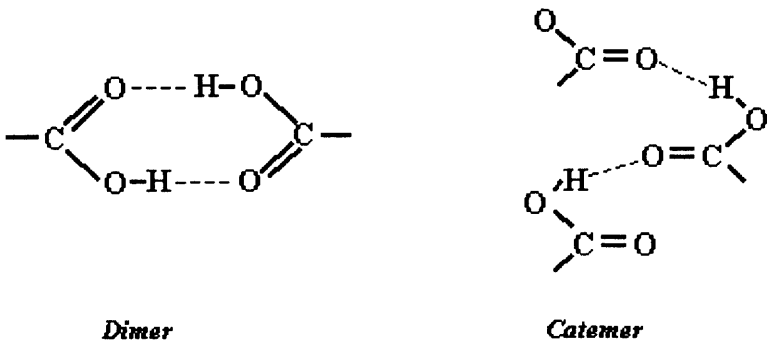


Fig. 1.1: Two typical hydrogen bond motifs in mono and di-carboxylic acids crystal structures.

b) Conformational Polymorphism

This particular kind of polymorphism arises when conformationally flexible molecules can be folded into different arrangements, in which they assume different conformations[24]. Different conformations would be expected to pack in different arrangements, so that sometimes the distinction between packing polymorphism and conformational polymorphism is artificial. An illustrative example is given by the piroxicam pivalate[31]. This kind of polymorphism has to be carefully distinguished from pseudo-polymorphism, which is the dynamic isomerism or tautomerism (eg. changes in covalent bonds), and is often treated as a distinct phenomenon from polymorphism[11].

Indeed, a range of phenomena related to polymorphism, are often all grouped under the term *pseudo-polymorphism.*

1) Solvates

These are crystalline solid adducts, obtained through crystallisation from solvent, which contain solvent molecules within the crystal structure in either stochiometric or no stochiometric proportions[24].

2) Hydrates are a particular class of solvates, in which the molecular solvent incorporated is water[24]. Because of its small size the water molecule can easily fill structural voids and due to its capability in forming multidirectional hydrogen bonds, it is ideal in linking drugs molecules in stable crystal structures.

3) Dynamic Isomerism or Tautomerism

This kind of phenomenon is quite rare, but a well investigated example is 2-amino-3-hydroxy-6-phenylazopyridine[32]. Often the tautomerism implies different properties for the two forms, such as different colours and different packing arrangements of the hydrogen bond networks. The phenomenon is related to the problem of definition of polymorphism. Dynamic isomers are considered chemically different species, but it is not always easy to make a distinction between geometric isomers species and conformationally different species. What essentially distinguishes isomerism and polymorphism is that the first exists either in the solid or in the molten phase and the two isomers are in equilibrium over a wide range of temperatures. On the contrary, in the case of polymorphism, the equilibrium between two different forms can occur only in the solid state, but in the molten state the difference disappears. Moreover at any particular temperature only one polymorph is stable, except at the transition point where two polymorphs are in equilibrium.

1.1.2 Properties and application of polymorphism in the organic solid state

The polymorphs of a certain compound have different crystallographic structures, different unit cell parameters and different crystal packing. Consequently they have different physical properties, including different thermodynamic, kinetic, mechanical, optical and surface properties. These different properties are important for the applications of organic solid materials in different fields. For example the different structures of polymorphic forms of a certain drug affect its solubility and rate of dissolution, strongly influencing the bioavailability of the compound. Hence it is necessary to have a proper strategy to control the formation and the production of a certain pharmaceutical compound with reliable and robust processes. The control of polymorphism is very important also in the field of explosives, in which the wrong polymorph can have a different sensitivity to detonation.

For molecular organic materials, the relationship between structural and property relations can be divided in two general categories. The first involves properties due to strong interactions between neighbouring molecules; the second concerns properties related to differences in the molecular structure, in general in the molecular conformation.

a) Electrical Properties: molecular organic materials are traditionally considered electrical insulators. In the 1970s some organic materials, crystals of the π molecular complex, such as tetrathiofulvalene/tetracyano-p-quinodimethane (TTF/TCNQ)[33] were discovered, which show metallic conductivity. These complexes crystallise with segregated stacks of molecules along the same crystallographic axis, each stack containing only one type of molecule as represented in figure 1.2

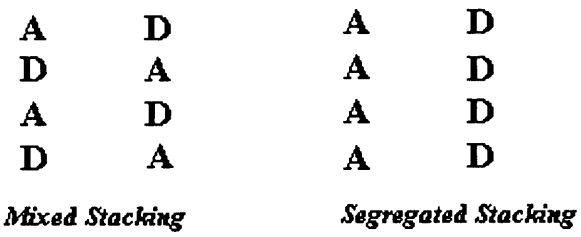


Fig. 1.2: Schematic representation of the stacking structure of the electronic conducting organic materials. A is the acceptor and D the donor.

This kind of stacking is a necessary condition for the electrical conductivity in these materials, which usually show the properties of polymorphism, with the mixed-stack form being the thermodynamically most stable. This latter one is also semi-conducting one, while the one structurally characterised by segregated stacks is metallic.

b) Magnetic Properties: another interesting class is the organic magnetic materials, for which different polymorphs show differ magnetic behaviour as in the case of the azaadamantane derivative dupeyredioxyl[34] and the tetracyanoethylene based compounds[35]. A design requirement for molecular magnetic materials is the presence of unpaired electrons on the molecular species. The magnetic behaviour is determined by the nature of the interactions among the electrons[36,37]. Changes in the molecular environment among different polymorphic forms can give changes in magnetic behaviour.

c) Photoelectrical Properties

There are several organic materials, which have applications in photovoltaicity and photoconductivity. Many are also used as pigments and dyes. They have a very important common feature: a direct relationship between crystalline structure and photoconductivity has not been firmly established, because of difficulties due to the presence of impurities and crystal size. Nevertheless the importance of a precise and specific stacking arrangement has been recognised[38]. Photoconductivity is a solid state phenomenon, which requires strong intermolecular interactions, short interplanar distances (up to ~ 3.5Å) and a large charge mobility upon excitation.

d) Non-linear optical properties: a difference in the stacking of the molecules can affect the refractive index of the crystal, which arises from the vibrating interactions of the electrons along different spatial directions (phonons). Another optical property of a crystal is the capability of generating second harmonic (S.H.G.) waves, which is used to double the frequency of electro-magnetic radiation. S.H.G. is strictly related to *non*-centrosymmetric space groups of a crystal. Nevertheless there are other important factors contributing to S.H.G. behaviour. A number of studies[39]demonstrated the importance of the space group symmetry and of details of packing in determining the non liner optical activity of molecular crystals.

e) Chromoisomerism, Photocromism , Thermochromism and Triboluminescence:

Chromogenic effects, regarding changes in colour of a certain substance, under structural or internal environment changes, are sometimes due to subtle structural changes in the crystal of the compound and in other cases are caused by changes in the molecular structure of the compound. *Photochromism* is the formation, destruction or the change of the colour of a certain compound due to light absorption and is present in certain compounds[40]. *Thermochromism* (change of colour with temperature) is attributed to proton transfer mechanism along molecular chains[41]. Change in colour due to an external

strain or deformation of the material (*mechanochronism*) has also been observed, together with *triboluminescence* behaviour, which consists of the generation of light by the material undergoing friction.

f) Mechanical Properties: the density and the molecular volume of a polymorph depend on the closeness of the molecules to each other. At the same time the stacking of the molecules can lead to particular interactions between the molecules, which determine the mechanical properties of the crystal. In the case of drugs, differences in stacking of the molecules can drastically change the solubility of the crystal, which can have serious consequences in possible use of the pharmaceutical.

1.2 Thermodynamic *vs* Kinetic Factors

1.2.1 Thermodynamics

Thermodynamically, a certain molecular system can crystallise if there is an overall decrease in its free energy. For a particular compound, the crystals likely to appear are those having the lowest free energy at a certain temperature, which is the essential thermodynamic condition for a stable crystal structure. Often this is counteracted by kinetic factors, which encourage the crystallisation as quickly as possible[42].

The free energy determines the relative thermodynamic stability of two polymorphic structures. It is usual to consider the Gibbs free energy for a solid at a certain pressure (P) and temperature (T). It can be written as:

$$G = H - TS \quad , \qquad [1.1]$$

where H is the enthalpy, S is the entropy of the crystal and T the absolute temperature. At $T=0K$ (absolute zero), if we neglect the zero-point energy contribution (which usually is very small), the entropic contribution to the free energy vanishes and the free energy corresponds to the lattice energy of the crystal.

If we consider a generic polymorphic system ("dimorphic system"), consisting of two phases (form α and form β), equation [1.1] involves a difference of energies and it can be written as[43]:

$$\Delta G = \Delta H - T\Delta S \quad , \qquad [1.2]$$

where

$$\Delta G = G^{\alpha} - G^{\beta} \quad ,$$

$$\Delta H = H^{\alpha} - H^{\beta} \quad ,$$

$$\Delta S = S^{\alpha} - S^{\beta} \quad . \qquad [1.3]$$

A general crystal structure is stable at a certain temperature T and pressure P, when:

$$[\Delta G]_{T,P} = 0 \quad , \qquad [1.4]$$

where G is the Gibbs free energy of the crystal.

For a dimorphic system, given the thermodynamic definition of G in [1.2], the condition [1.4] can be re-written as:

$$[H\text{-}TS]_{\alpha} < [H\text{-}TS]_{\beta} \quad , \qquad [1.5]$$

with H and S indicating respectively the enthalpy and the entropy of the system.

We can split the entropic contribution into two terms: the first is the configurational entropy, arising from disorder or from different molar volumes together with rotational contributions and the second is vibrational entropy, given by both intra and inter molecular vibrations. The entropy differences between two phases are often small. Hence we neglect the entropic contribution, with equation [1.5] reducing to:

$$H_{\beta} > H_{\alpha} \quad . \qquad [1.6]$$

The enthalpy of a crystal is directly related to its structure, and to the intermolecular interactions. Neglecting thermal expansion and the intra-molecular contribution we can identify the enthalpy H with the internal energy of the system, which is the lattice energy U_{latt}. This quantity is strongly determined by the intermolecular forces between the molecules in the structure. Hence the relative thermodynamic stability of the polymorphs is generally related to the evaluation of the lattice energy of the system.

As a simple approximation, the lattice energy can be considered equivalent to the enthalpy of sublimation of the crystal[44].

$$\Delta H_{subl} \sim U_{latt} \quad . \qquad [1.7]$$

The enthalpy of sublimation of a crystal can be expressed as[45]:

$$\Delta H_{subl} = H(\text{vapour})\text{-}H(\text{crystal}) = \Delta H(\text{condens}) + \Delta H(\text{conformation}) \quad . \qquad [1.8]$$

The first term refers to the condensation energy of a molecule, which has the conformation found in the crystal, and the second refers to the difference between the enthalpy of the molecular conformation in the crystal and that in the gas phase. The evaluation of the enthalpy of sublimation for a certain compound is not an easy procedure and is usually derived by the *Clausius-Clapeyron* equation:

$$\ln\left(\frac{P_2}{P_1}\right) = \frac{\Delta H_{subl}}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \quad , \qquad [1.9]$$

where R is the ideal gas constant and P is the vapour pressure of the solid. The latter is usually very small and its measurement is critical for a good evaluation of the sublimation enthalpy.

The enthalpy of condensation ΔH(condens) is given by:

$$\Delta H(\text{condens}) = \Delta U(\text{condens}) + P\Delta V(\text{condens}) + \Delta U(\text{kin}) \quad , \qquad [1.10]$$

where ΔU(condens) is the difference between the internal energy of crystal and the internal energy of the gas, whilst ΔV(condens) ~ V(vapour). Taking into account the molecular motion in the crystal and in the vapour phase, the kinetic energy for the vapour is ~ $3RT$ whilst the kinetic and potential energy for the crystal ~ 6RT (equipartition theorem). The new expression of the enthalpy of sublimation with respect to the crystal lattice energy is[43]:

$$\Delta H_{subl} = \Delta H(\text{condens}) = \Delta U(\text{condens}) + P\Delta V + (+3RT - 6RT) = \Delta U(\text{condens}) - 2RT \quad . \qquad [1.11]$$

In this expression we neglect the ΔH(conformation) contribution, which is zero for rigid molecules. For flexible molecules ΔH(conformation) can become important and hence must be taken in account.

Gavezzoti and Filippini compared several properties of different polymorphic crystals[46] at room temperature. The differences in densities ΔD among polymorphs are usually less then 0.5% and typical values of lattice energies U_{latt} are between 100 and 200 kJ mol^{-1}, with the energy difference between the two polymorphs having a limit of 25 kJ mol^{-1}. Differences in lattice-vibrational entropies are also very small.

Thermodynamically, the system can be classified in two ways, depending on the kind of transformation to which is subjected. This characteristic strictly depends on the relative difference of the entropy of the two phases, given $\Delta S \neq 0$:

- *Monotropic systems*: at all temperatures below the melting point of the crystal (T_m), one polymorph (i.e. form α) is more stable than the other one (i.e. form β). This is illustrated in the following diagram of the energies (figure 1.3).

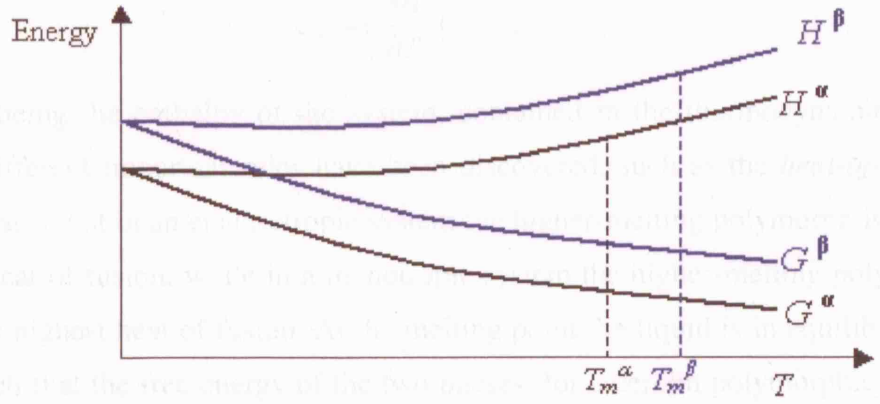


Fig. 1.3: Energy diagram (Free Energy and Enthalpy) of a *monotropic* polymorphic system, consisting of two phases α and β[19].

- *Enantiotropic system*: at a certain temperature (T_t), below the melting temperatures of the two phases, a phase transition (i.e. from form α to form β) can be observed, as shown in the energy diagram in figure 1.4.

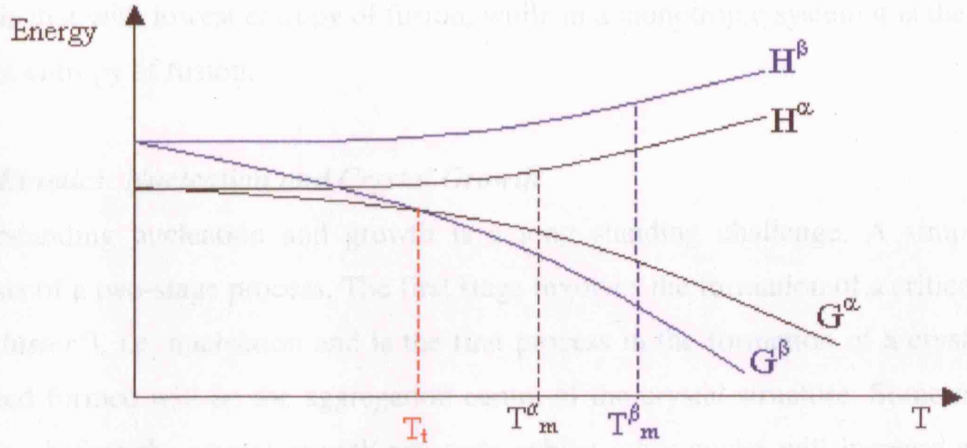


Fig. 1.4: Energy diagram (Free Energy and Enthalpy) of a *enantiotropic* polymorphic system, consisting of two phases α and β[19].

Before the temperature of phase transition T_t, form α is more stable than form β, because the enthalpy is still the dominating factor in the Gibbs free energy. At temperatures above the T_t, the entropic factor leads to the β phase, which is more stable than the form α ($G_\alpha < G_\beta$). The specific heat at constant pressure for a certain crystal is thermodynamically defined by the relationship:

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p \quad , \qquad [1.12]$$

with H being the enthalpy of the system, contained in the thermodynamic relationship [1.1]. Different important rules have been discovered, such as the *heat-of-fusion* rule[47], which states that in an enantiotropic system the higher-melting polymorph is the one with lowest heat of fusion, while in a monotropic system the higher-melting polymorph is the one with highest heat of fusion. At the melting point the liquid is in equilibrium with the solid such that the free energy of the two phases, for a certain polymorphic form, are the same:

$$\Delta G = 0 \quad . \qquad [1.13]$$

From the relationship [1.2], within this equilibrium condition at the melting point we can derive the *entropy-of-fusion* ΔS_f rule[48]:

$$\Delta S_f = \frac{\Delta H_f}{T_f} \quad . \qquad [1.14]$$

Between two enantiotropically related polymorphs, the form with the highest melting point is that with lowest entropy of fusion, while in a monotropic system it is the one with highest entropy of fusion.

1.2.2 Kinetics: Nucleation and Crystal Growth

Understanding nucleation and growth is a long standing challenge. A simple model consists of a two-stage process. The first stage involves the formation of a critical nucleus (or "*cluster*"), i.e. nucleation and is the first process in the formation of a crystal: hence the seed formed will be the aggregation centre of the crystal structure. Some nuclei can collapse before the crystal growth proceeds, whilst other nuclei will increase and grow. Finally other nuclei continue to be present, but their growth has stopped. In general we can distinguish between two types of nucleation processes[49]:

- *Primary nucleation*, which occurs for systems not containing crystalline matter.
- *Secondary nucleation*, which occurs in the presence of crystals of the solutes, in a supersaturated solution.

Furthermore, primary nucleation can be classified in two different processes:

- *Homogeneous nucleation*: this is a spontaneous process
- *Heterogeneous nucleation*: this is a process, induced by foreign particles or seeds.

This classification is briefly illustrated in the following scheme:

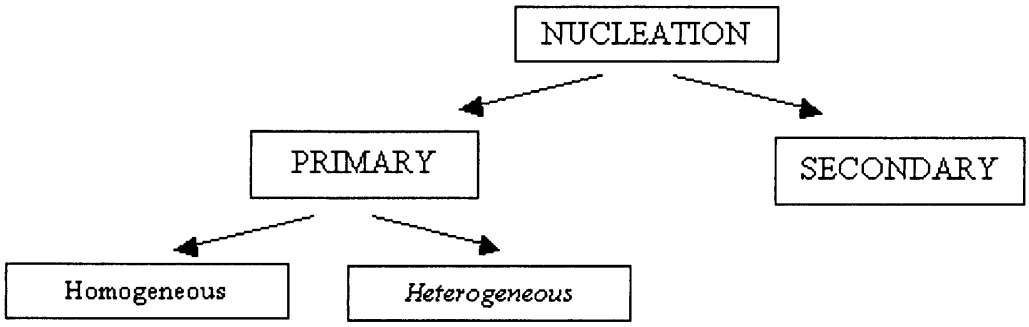


Fig. 1.5: Classification of the different types of nucleation processes in crystal formation[19].

To give a clue to the competitive role played by kinetics versus thermodynamics in crystal formation, we illustrate a possible mechanism of a homogeneous nucleation, proposed by Mullin[49]. It essentially suggests the existence of a bimolecular addition sequence to generate a crystalline nucleus. The mechanism is step-wise and leads to the formation of a critical "cluster", through a sequence of reactions in competitive equilibrium, as shown in this scheme:

$$\begin{aligned} Mol + Mol &\rightleftharpoons Mol_2 \\ Mol_2 + Mol &\rightleftharpoons Mol_3 \\ \ldots\ldots\ldots\ldots\ldots &\quad \ldots\ldots \\ Mol_{n-1} + Mol &\rightleftharpoons Mol_n \end{aligned} \qquad [1.15]$$

where *Mol* indicate a generic molecule.

Etter[50] extended this model, describing the clusters of molecules as aggregates, controlled by the same intermolecular interaction. An aggregate is an ensemble of molecules, whose centres of mass librate around an equilibrium position[44]. The aggregates are formed by the interaction of the molecules with each others, driven by the internal potential energy of the system. The latter acts against the kinetics random diffusion motion of the molecules. From a thermodynamic point of view the aggregate is in a potential energy valley , which cannot be disrupted by the librational motion. It has a recognisable pseudo-crystalline structure. In this model the aggregation/disaggregation step can be formally written as[44]:

$$[n_i Mol_i : n_j Mol_j] + Mol_i \leftrightarrow [(n_i+1) Mol_i : n_j Mol_j] \quad , \qquad [1.16]$$

where Mol_i and Mol_j are different molecular species, forming the aggregate, n_i and n_j are their respective number in the aggregate (which is represented by the square brackets).

The essential condition for the formation of a cluster or aggregate to precipitate from a solution is the *supersaturation* of the solution[51]. This term refers to a solution that

contains the dissolute material in a quantity, which is more than the one that could be dissolved by the solvent under equilibrium conditions.

The theoretical model of homogeneous nucleation establishes that the formation of a cluster is an energy activated process, which involves an energy barrier to be overcome, corresponding to the maximum free energy ΔG_c, in order to proceed to the formation of the crystalline nucleus. The size, corresponding to the maximum of the free energy, is the critical size n_c of the nucleus. Clusters at critical size are called critical nuclei. For sizes less than this critical value we will observe dissolution of the forming clusters, while for sizes bigger than this critical value we will observe growth of the cluster. The chance of forming nuclei of this critical size depends on the height of the free energy barrier relative to thermal energy k_BT. The inverse process, the dissolution of a cluster, depends as well on the size of the aggregate.

The rate of nucleation J is another important factor in the formation of a crystal. It is defined as the total rate at which atoms join *critical nuclei*, making the cluster stable.

1.2.3 Nucleation in polymorphic systems

The crystallisation of a compound, which has more than one solid phase, often follows the so called *Ostwald's rule*: "an unstable system does not necessarily transform into a more stable state, but into one which most resembles the former, i.e. into another transient state whose formation from the original is accompanied by the smallest loss of free energy"[49]. Hence the crystallising system goes from the supersaturated state to the equilibrium, through different stages, which represent the smallest possible change in free energy. Thus a crystallising polymorphic system would move through all the possible different stages, and structures, before crystals of the thermodynamically most stable structure appear. Hence in a polymorphic system with multiple equilibria, the equilibrium would shift in favour of one polymorphic form, once the nucleation occurs (figure 1.6).

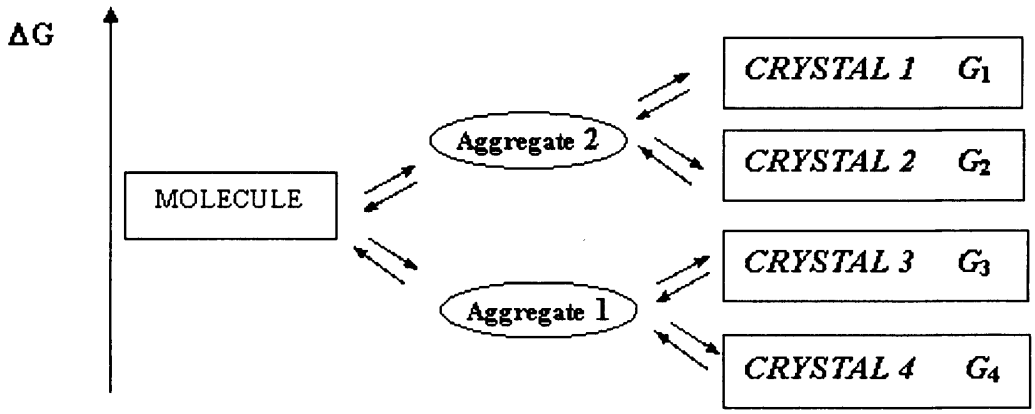


Fig. 1.6: scheme of the nucleation process, involving the formation of two different aggregates. $\Delta H_4 < \Delta H_3 << \Delta H_2 < \Delta H_1$.

Although crystal 4 is the most thermodynamically stable structure (because of the lowest enthalpy), if we suppose that aggregate 2 is more easily achievable than aggregate 1 (in terms of faster nucleation and crystal growth), we obtain crystal form 2 at the end of the crystallisation process. This structure is less preferred thermodynamically, but kinetics determines its formation. Form 4 is thermodynamically more stable than form 2 ($G_4 < G_2$), so that in principle we would expect the crystallisation of form 4. The key step in the crystallisation of the two polymorphic forms from the supersaturated solution is the formation of nuclei and aggregates, which have a specific size and interface with the solution. These are critical factors in the kinetics of the crystallisation.

1.2.4 Disappearing Polymorphism

In the organic solid state the experimental structure, obtained under certain conditions of crystallisation, is frequently not the most stable. Sometimes the appearance of a new polymorph can change the environment, in which the polymorph was found and the previously observed crystal cannot be crystallised anymore. The new polymorph is likely to be thermodynamically and energetically more stable and the possible presence of seeds of the new form can change the environment, in which the previous crystal was grown. In this way the re-formation of the old, less stable structure is irreversibly prevented. A more stable structure replaces the previously observed crystal. This phenomenon is known as "*Disappearing Polymorphism*"[52]. Many cases are reported in the literature about crystal structures, whose forms, previously produced over a long period of time (years), suddenly could not be crystallised anymore after a new more stable polymorph had been discovered[53]. Usually the polymorphs first obtained, which then disappear, are

the metastable forms of the compound, which are later replaced by the thermodynamically most stable polymorph, in agreement with *Ostwald's Rule* (see section *1.2.3*). Hence the disappearing form is often the metastable one. Often intentional or unintentional *seeding* plays an important role in the phenomenon of disappearing of the metastable structure[54]. The presence of microscopic seeds of the most stable structure in the crystallisation environment can come into contact with the metastable form and act as a nucleating agent of the stable phase. The most illustrative, dramatic example is that of the antiviral ritonavir[55], a drug used for the treatment of the Acquired Immunodeficient Syndrome (AIDS). A critical issue concerns the possibility and necessity of controlling the metastable form, preventing its disappearance at the expense of the most stable structure. The formation of the metastable form can be led, by introducing specific additives, impurities or reaction by-products[56] in the crystal, which can inhibit the nucleation and growth of the most stable form at the expense of that yielding the metastable phase, as in the case of sulfamerzine[57] and sulfathiazole[26].

In conclusion, in polymorphism, different conditions affect crystal formation and it is very difficult to pinpoint the key factor in the process. In crystal formation, kinetics plays a fundamental role, determining the morphology of the crystal and the crystal growth. Important kinetic factors are the aggregation of first nuclei and the nucleation rate. These factors often counteract the thermodynamic ones. A detailed discussion of all these different factors is given by Kitamura[58].

1.3 Crystal Engineering and Crystal Structure Prediction (CSP)

Crystal structure prediction remains a major challenge. In 1988 John Maddox wrote[59] that "*one of the continuous scandals in the physical sciences is that it remains in general impossible to predict the crystal structure of even the simplest crystalline solids*". By "crystal structure prediction" we mean the ability to predict in advance a crystal structure of a certain compound, simply starting from atomic or molecular contents[20]. In the last 20 years major improvements have been achieved in the investigation of the solid state, using both experimental and computational methods. In particular the latter, has acquired increasing importance in the investigation of the solid state, thanks to growth in computer power and improved methods. Crystal structure prediction embraces both the inorganic as well as the organic world. Here we focus on the crystal structure prediction of the organic solid state, in which attention has increased rapidly in the recent years owing to its fundamental interest and commercial importance[19,60]. The state of matter in this research

field is well summarised in two papers published almost 10 years from each other, both entitled: *"Are crystal structures predictable?"*[61,62]. In some cases the answer to this question is "*No*", but much progress have been achieved[63].

1.3.1 Crystal Packing

The structure of an organic molecule can be described in terms of bonds between atoms, and angles and torsions between the molecular planes, which are important in determining the molecular structure. In addition, a molecule also has shape and size, not only in terms of chemical bonding backbone, but also in terms of surface and volume. Indeed, effective radii can be assigned to atoms in the molecule[64]. These radii can be determined by diffraction experiments and, in conjunction with the atomic positions, they can be used to define a space-filling model of a molecule. A molecule has a series of properties, which include molecular weight (M_W), molecular volume (V_M), molecular density (D_M) and molecular free surface area. The latter is a good indicator of molecular size[65].

On the other hand crystals can be seen as the lattice vectors applied successively to the asymmetric unit cell, which is the smallest unit, whose translation in space defines the crystal structure. The number of molecules (the component) in the asymmetric unit cell Z′ is an important parameter in organic molecular crystals, as well as the packing coefficient (*P.C.*) of a crystal (which is a very good index of packing efficiency of the molecules into a crystal). The latter was introduced by Kitaigorodskii[64,66], who was the first to realise that the molecules in molecular crystals tend to pack so as to form a close packed structure, as known for ionic crystals. The *P.C.* is defined as the ratio between molecular volume (V_M) and cell volume (V_C), by the relationship:

$$P.C. = \frac{Z \cdot V_M}{V_C} \quad , \qquad [1.17]$$

where Z is the total number of molecules in the unit cell.

In molecular crystals the average value of the packing coefficient is around 0.7, not far from that for spheres or ellipsoids. Molecules with irregular shape, with many concave or convexes surfaces, may nevertheless form very compact structures, if they are in very close contact with each other and occupy the space in the crystal uniformly. For crystals of large molecules, which necessarily crystallise leaving voids in the crystal structure, the closed-packing rule is often maintained by the presence of solvent, incorporated in the structural voids, with consequent formation of solvates/hydrates.

Molecules pack together in a crystal owing to the intermolecular forces between them. The strength of these forces can be very different, from species to species, which seems to have very limited influence on the packing coefficient of the crystal: not even directional intermolecular forces, like hydrogen-bond or specific electrostatic interactions can violate this general closed-packing rule. Kitaigorodskii theory[64,67] indicated specific symmetry operators as the sole possible operators permitting a close-packing structure:

- Inversion centre
- Twofold screw axis
- Glide plane
- Translation.

As consequence only a limited number of space groups, arising from the combination of these symmetry elements are observed in the molecular organic solid state[44,68]. This insight was demonstrated by a survey of the CSD, which shows the 90% of all determined crystal structures belong to close-packing space groups: *P1*, *P-1*, $P2_1$, $P2_1/c$, $P2_12_12_1$, $P2_12_12$, $Pca2_1$, *Pna21*, *Pbcn*, *Pbca*, *Cc*, *C2*, and *C2/c*.

It might be thought that a higher packing coefficient could result from stronger attractive forces between interacting molecules in the crystal, irrespective of the shape. The result of an analysis of crystals of different organic compounds, with a variety of substituent groups, which generate stronger intermolecular forces (for example polar groups in the place of hydrogens), indicates no modification of the packing coefficient. It is not obvious, on the contrary, that crystals with low packing coefficient have low packing energies, since the interplay between molecular shape and interatomic potentials is too complicated to be analysed in such simple terms. Still, molecules, which are unable to close-pack, may be not crystalline at all.

Overall it is important to underline that close-packing is a necessary, but not sufficient condition, for an organic crystal to form[44]. A computational crystal structure search can generate hundreds of closed-packed structures.

1.3.2 Crystal Packing Motifs and Crystal Engineering

The crystal packing motif describes the packing of molecular units inside a crystal structure[69]. The way the molecules assemble in a crystal is governed by the intermolecular interactions between them, whether these are ionic, hydrogen bonds or weak *non*-bonded interactions. The intermolecular interactions are in turn governed by the relative position of the molecules in the crystal.

Desiraju[70] gives the definition of Crystal Engineering as "*the understanding of intermolecular interactions in the context of the crystal packing and in the utilisation of such understanding in design new solids with desirable physical and chemical properties*". The crystal engineering view on the organic solid state considers a crystal as a supramolecular entity, formed by specific intermolecular arrangements, given by the interactions of the constituent molecules during the crystallisation process. The concept of the "supramolecular synthon" is often considered in crystal engineering[71] and is defined as the pre-stable arrangement formed in the first stages of the crystallisation process, which already shows the structural motif of the future crystal. Hence an understanding of how specific commonly occurring molecular functional groups or intermolecular motifs are likely to interact in the crystalline environment is important. A very useful instrument, which allows statistical studies of intermolecular interactions of the molecules over classes of structures, is the CSD[72].

One of the main motifs, well recognised as a protagonist in organic crystal chemistry, is the hydrogen bond X-H···Y (X,Y= N, O), even if there is not a specific theory developed for it. Typical H···Y distances are around 2Å or below. This interaction is very common in alcohols, carboxylic acids, amides and whenever possible this interaction will form. There are exceptions, which are represented for instance by the case of the alloxan crystal structure, on which a detailed investigation of the intermolecular forces was performed[73]. The result showed that C=O···C=O interactions are competitive with a very weak C=O···H-N hydrogen bond, which is probably due to the unusual bonding environment of the C=O and N-H groups in the molecules. Other crystal structures of this compound, containing hydrogen bond interactions in the form of *dimers* (see section 1.1.1), were predicted, but not thermodynamically as stable as the experimental structure[73]. Many studies were performed in the area of carboxylic acids[74] and primary[75] and substituted amides[76]. For monocarboxylic acids, typical motifs are cyclic hydrogen-bond pairs, but *chains* of molecules, linked by linear hydrogen-bond interaction O-H···O are also common. In amides ($R\text{-}CONH_2$), the molecules tend to crystallise via hydrogen bond rings linked by pair of C=O···H-N interactions, in different ways, given by the translation of the N-H···O bonds along glide planes, or twofold screw axes. Other specific hydrogen bond interactions were found, like the C-H···O interaction[77], which was established to be significant in crystal chemistry and its energetic implications are mainly founded on electrostatic grounds[78]. Another special bonding effect is that arising from flat parallel aromatic rings, observed in crystal structures of benzene and polyaromatics[79].

Another important statistical study using the CSD is that concerning the number of molecules in the asymmetric unit cell (Z′)[80]. In molecular crystals, the most common value is Z′=1 and only a few structures have Z′ > 1. These are mainly nucleotides, nucleosides and steroids. Quite often a polymorph has a Z′ > 1 in a special subgroup of the partner polymorph with Z′ =1. Although with certain exceptions it is generally observed that some of the different forms with Z′ > 1 in the asymmetric unit are correlated by pseudosymmetry. A typical value is Z′=2 (especially in salts and organometallic crystal structures), but also Z′=3 and Z′= 4 are observed (e.g. pyridine[81]). Values of Z′ > 4 are quite rare and associated typically with puzzling structures. Other typical values are Z′ = ½, which usually occur for molecules with high symmetry, which occupy one special position (e.g. a mirror plane) in the crystalline unit cell. An example is provided by one of the polymorphs of the terephthalic acid[82].

1.3.3 Crystal structure prediction procedure

Computational crystal structure prediction is essentially based on the search for the crystal structure with the lowest lattice energy U_{latt}. Following the thermodynamic approach, this method is complementary to the crystal engineering approach and it gives more insights about the relative stability of the polymorphs by calculating the lattice energies of the system. Thus crystal structure prediction has considerable potential to aid crystal engineering[83].

Starting from a simple molecular structure, the methods try to predict the cell parameters, space group and molecular location and orientation in the unit cell of all the possible stable structures. In the case of flexible molecules, usually many possible energy stable conformations usually have to be considered.

Molecular clusters are constructed by considering the symmetry operators of the most common space groups in the organic solid state. The unit cells are accepted on the basis of the closed-packing rule, for which they have to have a packing coefficient of ~ 0.7. Classification of all the preliminary calculated lattice energies is used, in order to remove structures with too strong repulsive contacts. At this level, a model based solely on repulsive intermolecular forces is used.

For each of the possible structures generated, a minimisation of the internal lattice energy is performed, using a full interatomic potential model. This point is absolutely crucial and is often the key to the success of a prediction. The efficacy of the prediction depends on the accuracy of the potential model. This approach requires also many other assumptions

about the molecular structure, which make the model of the intermolecular forces very dependent on the type of molecule investigated[83]. After energy minimisation, the structures generated by the prediction are typically thousands and many are redundant. Hence a *clustering* procedure is necessary, which eliminates the structures, which are essentially the same. This is not always a trivial step, because very similar structures can have different unit cells and even different space groups. Having obtained the final different stable minimised structures, the experimental structure (if resulting in the thermodynamically stable polymorph) should in principle match the global minimum structure, and metastable forms should correspond to local minima.

Typically the search method produces thousands of possible structures and after clustering and lattice energy minimisations still hundreds structures are present and around tens are in the range of 10kJ mol^{-1} from the global minimum[83].

Understanding whether other minima represent effective possible polymorphs is a very delicate issue and often remains without a definitive answer. Nevertheless calculation of elastic constants and phonon frequencies can be useful in investigating more deeply the mechanical and energetic stability of the structures[84]. Further calculations of the in *vacuo* morphology and growth rates of the structures can sometimes help in evaluating the possibility of a certain structure to be a candidate polymorph[83]. A general scheme of the crystal structure prediction procedure is shown in figure 1.7 and the methodology is discussed in more detailed in chapter 3.

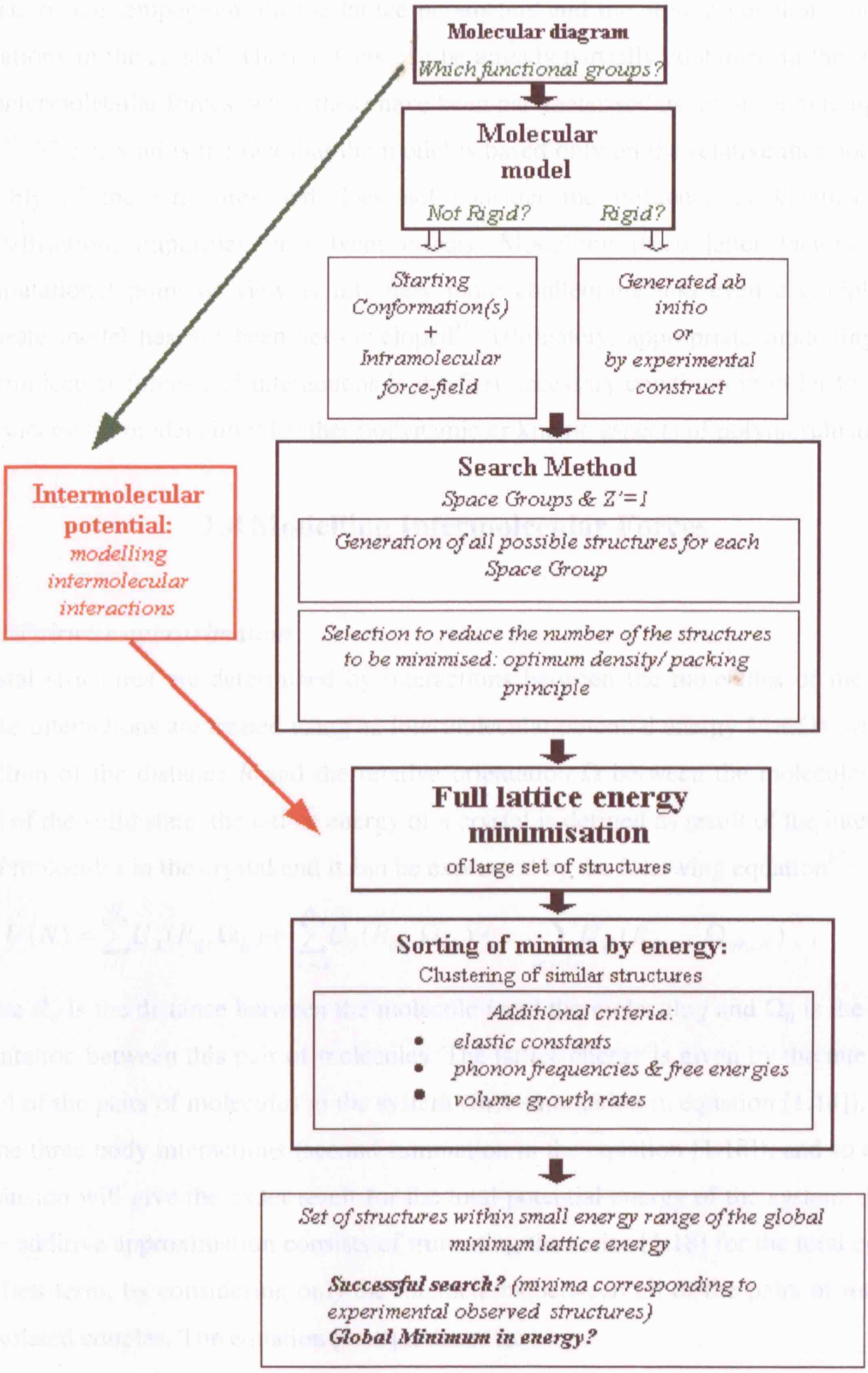


Fig. 1.7: Flow-chart scheme of a typical computational crystal structure prediction. The molecule-dependent choices are shown in green.

There are two important limits to this approach. The first is the neglect of thermal contributions[85]. All the predictions are effectively at 0K and do not take in account the

effects of the temperature on the lattice parameters and the atomic positions and lattice vibrations in the crystal. These effects can be already partially contained in the model of the intermolecular forces, when these have been parameterised using of room temperature data[85]. The second is the fact that the model is based only on the relative thermodynamic stability of the structures and does not consider the influence of kinetics on the crystallisation, impurities or solvent effects. Modelling these latter factors from a computational point of view is infinitely more challenging and even a complete and accurate model has not been yet developed[86]. Ultimately, appropriate modelling of the intermolecular forces and interactions is the first necessary condition in order to develop any successful model either for thermodynamic or kinetic aspects of polymorphism.

1.4 Modelling Intermolecular Forces

1.4.1 Pairwise approximation

Crystal structures are determined by interactions between the molecules of the crystal. These interactions are treated using an intermolecular potential energy $U(R,\Omega)$, which is a function of the distance R and the relative orientation Ω between the molecules. In the case of the solid state, the lattice energy of a crystal is defined as result of the interactions of N molecules in the crystal and it can be expressed by the following equation[87]:

$$U(N) = \sum_{i<j}^{N} U_2(R_{ij},\Omega_{ij}) + \sum_{i<j<k}^{N} U_3(R_{ijk},\Omega_{ijk}) + ... \sum_{i<j<k<...}^{N} U_N(R_{ijk...N},\Omega_{ijk...N}) \quad , \qquad [1.18]$$

where R_{ij} is the distance between the molecule i and the molecule j and Ω_{ij} is the relative orientation between this pair of molecules. The lattice energy is given by the interactions of all of the pairs of molecules in the system (first summation in equation [1.18]), plus all of the three body interactions (second summation in the equation [1.18]), and so on. This expansion will give the exact result for the total potential energy of the system. The pair wise additive approximation consists of truncating the series [1.18] for the total energy at the first term, by considering only the interactions between all of the pairs of molecules as isolated couples. The equation [1.18] reduces to:

$$U(N)_{pairwise} \sim \sum_{i<j}^{N} U_2(R_{ij},\Omega_{ij}) \quad . \qquad [1.19]$$

The error in this approximation depends strongly on the characteristics of the molecule, such as its charge distribution and on the kind of interactions present in the system.

The intermolecular interactions can be classified in two types:

- *long range interactions:* electrostatics, dispersion, induction (polarization).
- *short range interactions:* exchange-repulsion, charge transfer.

1.4.2 Long range interactions: Electrostatics

The electrostatic energy between a molecule A with a certain number n_A of electronic charges q_j and molecular volume V_j and a molecule B with n_B charges q_k molecular volume V_k can be expressed by the classical formula:

$$U_{el} = \frac{1}{4\pi\varepsilon_0} \sum_{j=1}^{n_A} \sum_{k \neq j=1}^{n_B} \frac{q_j q_k}{r_{jk}} \quad . \qquad [1.20]$$

The electrostatic contribution to the intermolecular forces is always strictly additive.

In a very simple model the electronic structure of the molecule can be represented by a set of point charges, located at nuclei which may be derived from an *ab-initio* quantum calculation using several techniques[88](e.g. Mulliken population analysis[89]), which calculates the atomic charges directly from the wave function, assuming that the overlap between two orbitals of two different atoms is shared equally[90]. This method does not reflect very well the different electronegativity of the individual elements and is efficient only in the case of small basis sets. With large basis sets (usually necessary for the evaluation of the electronic structure of organic molecules) it can give very unreasonable results[90]. In all these cases, the charges on the atoms do not have any significant physical meaning[88].

A more sophisticated method is to consider the electron densities ρ_j ($\boldsymbol{r}$) and ρ_k ($\boldsymbol{r}$) respectively of molecule A and B. The electrostatic energy can be calculated by considering the product of the electronic densities of the molecules:

$$U_{electr.} = \int_{AllSpace} \frac{\rho^A(\boldsymbol{r}_1) \cdot \rho^B(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} d^3 r_1 d^3 r_2 \quad . \qquad [1.21]$$

The electronic densities can be evaluated from *ab-initio* calculations. An alternative technique to evaluate the atomic charge model electron density of a molecule is to fit the electronic charges to the electrostatic potential (ESP) of the molecule[90]. The technique consists of:

- using the wave function of the molecule (obtained by the SCF procedure) to calculate the molecular electrostatic potential on a grid of points outside the molecule, in the regions where we want to represent its interaction with other molecules.

- fitting the atomic charges to this values of the potential in order to reproduce the molecular electrostatic potential. This gives the best possible model for a certain series of grid points and a certain choice of sites for the charges and a specific algorithm to do this is discussed in more detail in section 4.2.1.

The latter method assumes the charge distribution around each atom to be spherical so that the final distribution of the molecule is the sum of all the spherical point charges centred on each atom. Many different algorithms have been developed in fitting charges to the ESP[91,92]. This method gives much more accurate results in the field of organic molecules. Its disadvantage is that it depends strictly on the molecular conformation in the gas phase, so that for flexible molecules, the charges have to be re-calculated for every different conformation.

For neutral molecules, having *non*-spherical charge distributions, the use of atomic point charges alone does not always provide a good estimation of the molecular interactions. Although in some cases useful results may be obtained[93], in general, the assumption of spherical charge distribution around the atoms is limiting for molecules containing polar groups[93]. In fact, electrostatic interactions such as hydrogen bond, lone pair electrons and π bonds, given by molecular polar groups, such as N-H or C = O, are strongly directional in space and the latter are not well described using only a point charge model[93]. An improved approach consists of expanding the charge distribution of a molecule in a multipoles series (point charge, dipole, quadrupole, octupole and so on), which can better model the asymmetry in the molecular electronic structure.

The method consists of expanding the term $1/\boldsymbol{r}_{ij}$ in the expression of the electrostatic interaction, in a series, which is a function of the distance between the molecules $\boldsymbol{R}$. In this way, the electrostatic charge distribution will be divided into configurations with different symmetries, like point charges, dipoles, quadrupoles, octupoles, etc., as shown in figure 1.8. From a theoretical approach, considering the quantum Rayleigh-Schroedinger perturbation theory[94], the final expression for the electrostatic interaction between two molecules A and B is:

$$U_{electr.} = \sum_{l_1,l_2,k_1,k_2} \frac{(l_1 + l_2)!}{l_1! l_2!} \cdot Q_{l_1 k_1}{}^{A} \cdot Q_{l_2 k_2}{}^{B} \cdot \overline{S}^{k_1,k_2}{}_{l_1,l_2,l_1+l_2}(\Omega) \boldsymbol{R}^{-l_1-l_2-1} \quad . \quad [1.22]$$

In this expression $\overline{S}^{k_1,k_2}{}_{l_1,l_2,l_1+l_2}(\Omega)$ is a function of relative orientation of the molecules and $Q_{l_1 k_1}$ is the permanent multipole moment of the isolated molecule in a tensor notation, which represents the charge distribution of the molecule. The number l

and k are the eigenvalues of the angular momentum in quantum mechanics equation, so they are respectively analogous to the angular momentum and the magnetic momentum quantum numbers.

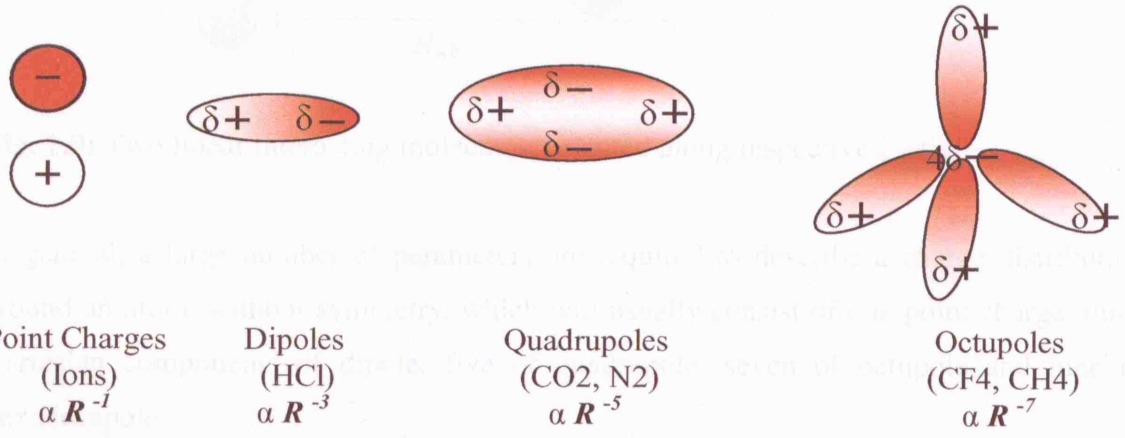


Fig. 1.8: Schematic representation of the multipoles expansion of a charge distribution. ***R*** is the distance between the molecules.

Usually the multipole expansion is used up to the R^{-5} term, which includes the quadrupole-octupole and hexadecapole-charge interactions[95]. For two linear interacting molecules (see figure 1.9) with axis directions defined by the unit vectors z_A and z_B and having a dipole ($\mu = Q_{10}$), a quadrupole ($\Theta = Q_{20}$) and an octupole ($\Omega = Q_{30}$), the electrostatic interaction energy has the following expression:

$$U_{electr}^{AB} = \frac{\mu_A \mu_B}{4\pi\varepsilon_0 R_{AB}^{3}} \cdot f_1(\theta_A,\theta_B,\phi) + \frac{\mu_A \Theta_B}{4\pi\varepsilon_0 R_{AB}^{4}} \cdot f_2(\theta_A,\theta_B,\phi) - \frac{\mu_B \Theta_A}{4\pi\varepsilon_0 R_{AB}^{4}} \cdot f_2(\theta_B,\theta_A,\phi) +$$
$$+ \frac{\Theta_A \Theta_B}{4\pi\varepsilon_0 R_{AB}^{5}} \cdot f_3(\theta_A,\theta_B,\phi) + \frac{\mu_A \Omega_B}{4\pi\varepsilon_0 R_{AB}^{5}} \cdot f_4(\theta_A,\theta_B,\phi) - \frac{\mu_B \Omega_A}{4\pi\varepsilon_0 R_{AB}^{5}} \cdot f_4(\theta_B,\theta_A,\phi) + ... \quad [1.23]$$

The functions $f_i(\theta_A,\theta_B,\phi)$ are functions of the relative molecular orientations, expressed by angular coordinates. The higher multipole moments represent the electrostatic potential arising from the non-spherical features in the atomic charge distribution such as lone pairs and π electrons.

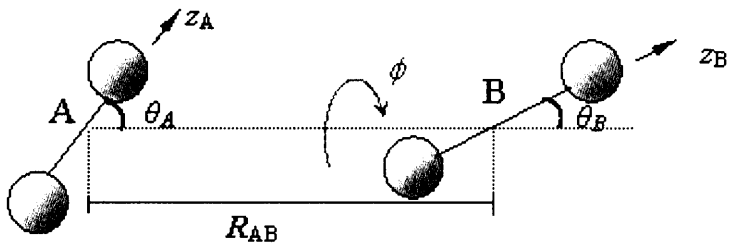
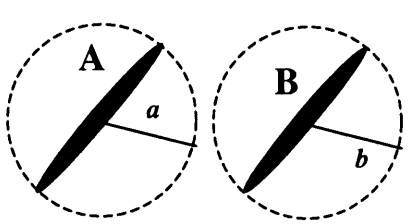


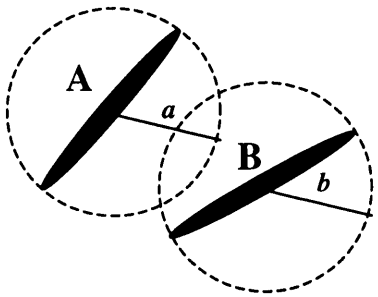
Fig. 1.9: Two linear interacting molecules, oriented along respective z axis.

In general, a large number of parameters are required to describe a charge distribution around an atom, without symmetry, which will usually consist of one point charge, three Cartesian components of dipole, five of quadrupole, seven of octupole and nine of hexadecapole.

In the *central multipole expansion*[94], the expansion is applied to the entire molecular charge distribution, so that the multipoles are centred on the molecular centre of mass. The multipole expansion is defined within a sphere of convergence having a certain radius. When considering two interacting molecules that are sufficiently far apart so that there is no overlap between the charge distributions, the central multipole expansion is only possible when the respective spheres of convergence of the multipole for the two molecules do not overlap, which makes its application invalid for certain relative orientations of non-spherical molecules (see figure 1.10).



Convergent Expansion for $U_{electrost.}^{AB}$



Divergent Expansion for $U_{electrost.}^{AB}$

Fig. 1.10: Molecular central multipole expansion convergence spheres and relative orientations in two interacting molecules.

This problem can be avoided using a *distributed multipole expansion*[94], which consists of using a set of multipoles at many sites usually the atoms in the molecule. The multipole expanded charge distribution is now the charge distribution on each atom of the molecule, and the sphere of convergence is defined through the van der Waals radius of

the atom. The multipole expansion of the atomic charge converges much more rapidly than the that of the charge distribution around the molecule[96]. This method requires a procedure to divide the molecular electronic charge assigning a certain amount of charge around each atom, which is then expanded in multipoles. The procedure used in this work is the Distributed Multipole Analysis (DMA) of Stone[97,98]. Usually the multipoles are "shifted" to the nearest atomic sites, for higher accuracy, since the process of "shifting" optimises the convergence of the DMA method. The use of multipole expansion is critical for compounds with particular directional electrostatic interactions. In the study of benzene dimers[96], the sole use of point charges (isotropic model) would predict binding above the centre of the aromatic ring, giving as result the parallel stacking of the molecules, while a careful distributed multipole analysis predicts a T-shaped or displaced parallel benzene dimer, optimising the π interactions between the benzene rings. In the study of the crystal structure of many other compounds, such as allopurinol[99], propane[93], other hydrogen bonded systems[100] and compound III of the CSD blindtest[101] in 1999 (which will be discussed in more detailed in chapter 5), much better results were obtained by using distributed multipoles instead of simple point charge models in the description of the electrostatic interaction.

1.4.3 Long range interactions: Induction (Polarization)

The induction energy arises from the distortion of the molecular charge distribution, due to the interaction with the field from the undistorted charge interaction of the other molecules. This contribution can become very important, in particular in crystals having strong electrostatic fields, i.e. ionics, and organics with polar groups in the molecule (figure 1.11).

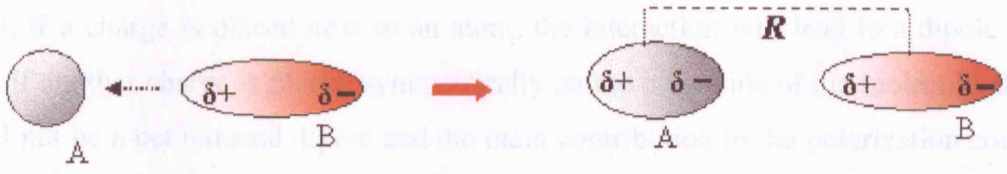


Fig. 1.11: Polarisation induction effect on a molecule, by the presence of a permanent polarised molecule.

In the Rayleigh-Schroedinger quantum perturbation theory[89], the expression of the inductive term for two molecules A and B, is given by the expression:

$$U_{induction}{}^{A} = -\sum_{n^A \neq 0} \frac{\langle 0^A 0^B \mid H' \mid n^A 0^B \rangle \cdot \langle n^A 0^B \mid H' \mid 0^A 0^B \rangle}{E_n{}^A - E_0{}^A} \quad , \qquad [1.24]$$

where H' is the perturbation hamiltonian of the system, given by the equation:

$$H' = \sum_{i<j} \frac{e_i{}^A e_j{}^B}{4\pi\varepsilon_0 r_{ij}} \quad . \qquad [1.25]$$

In equation [1.24] $|0^A\rangle$ and $|0^B\rangle$ are the ground states of the two molecules A and B and $|n^A\rangle$, $|n^B\rangle$ are the excited states of the two molecules.

There is another term, which is very similar to [1.24], but with the molecule B in the distorted, excited state $|0^A n^B\rangle$ and molecule A in the undistorted state $|0^A 0^B\rangle$. An important study by Mooij *et al*[102] attempted to evaluate the contribution of induction energy by fitting the polarisation part of a classical interatomic potential to *ab initio* data. The results obtain for ethanol dimers and trimers are satisfying and encouraging[102]. The classical part of the induction contribution can be calculated from the relationship:

$$U_{induction} = -\int_{All\ Space} \boldsymbol{\mu}\, dE = -\alpha \int_{All\ Space} \mathbf{E}\, dE \quad . \qquad [1.26]$$

In equation [1.26] $\boldsymbol{\mu}$ is the electric dipole moment created on generic molecule A, interacting with the other molecules. This dipole moment is generated by the presence of an average electric field **E,** created by the other molecules around the molecule A. α is the isotropic atomic polarizability. The nature of this interaction depends on the geometry of the molecules and on the electric field at the molecule. It can depend on the distance between the molecules as $\boldsymbol{R^{-4}}$ (distance dependence of the field generated by the spherical undistorted charge distribution of the molecule, which induces a dipole on adjacent molecules) or $\boldsymbol{R^{-6}}$ (distance dependence of the field generated by a molecule having an intrinsic dipole on adjacent molecules). Hence the interaction is very non-additive. In fact, if a charge is placed next to an atom, the interaction will lead to a dipole moment, but if another charge is placed symmetrically on the other side of the molecule, then there will not be a net induced dipole and the main contribution to the polarization comes from the quadrupole term. A *central distribution multipole expansion* can also be used to give an expression for the induction energy in terms of permanent multipoles and polarisabilities of the interacting molecules in the system. Using this method the calculated induction energy suffers from the same problems of convergency and accuracy, occurred in the treatment of the electrostatic interaction. Improvements are given by the use of a distributed polarisability model. Stone developed a model to treat the induction contribution, the Distributed Polarizability Analysis (DPA)[103] and its use

demonstrates the need to take into account this contribution in the crystal structure prediction of molecular solids. The model is too elaborate, being applicable only to very simple systems, and other more simple models[104] are not very accurate and transferable. Moreover, the *ab initio* calculation of the polarization is basis set dependent and is prohibitively expensive for large molecules. The usual practice is to ignore this contribution, given its expensive computational cost and its strong non-additive nature, and to assume the contribution to be adsorbed in the empirical parameters of the interatomic potentials.

1.4.4 Long range interactions: Dispersion

The dispersion contribution comes purely from a quantum mechanical effect and represents the stabilisation arising from the correlation of the charge fluctuations in the molecular distribution. In the Rayleigh-Schroedinger perturbation theory[89] for two interacting molecules, it is expressed by the term:

$$U_{dispersion} = -\sum_{\substack{n^A \neq 0 \\ n^B \neq 0}} \frac{\left|\langle 0^A 0^B \mid H' \mid n^A n^B \rangle\right|^2}{E_n{}^A - E_0{}^A + E_n{}^B - E_0{}^B} \quad , \qquad [1.27]$$

where H' is the perturbation Hamiltonian in the second order perturbation theory and the states $|0^A 0^B\rangle$ and $|n^A n^B\rangle$ represent respectively the ground state and the double excited state of the system, formed by the two molecules. The energy $E_n{}^A$-$E_0{}^A$ is the difference in energy of these states, whose dipole transition of the molecule A is represented by the term $\langle 0^A \mid \mu_A \mid n^A \rangle$.

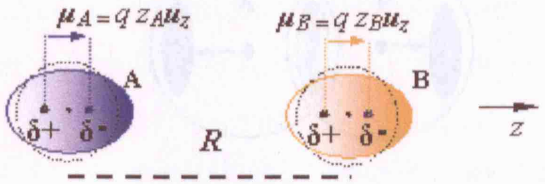


Fig. 1.12: Oscillating electronic charge in two adjacent atoms, their dipole moments and correlated electronic motions. q is the generic electronic charge of the molecule A and B.

Application of the multipolar expansion gives the dispersion energy between two spherical charge distributions (figure 1.12) as:

$$U_{dispersion} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - .... \quad . \qquad [1.28]$$

Usually the equation [1.28] is truncated at the first term.

This dispersive contribution is additive to the second order of perturbation theory, or in the classical expression [1.28] if we truncate the series at the first term. Considering higher order contributions, such as three-body contribution (Axilrod-Teller-Muto term[105]), added for better accuracy in a few particular cases (like argon[106]), the dispersion becomes non additive and the sign of the contribution (negative or positive) depends on the spatial orientation of the molecules. The C_n constants can be obtained from spectroscopic data or from *ab initio* calculations, but the latter are obtainable only for small molecules[102]. Accurate values from *ab initio* calculation require very high quality wave functions[107].

Density Functional (DFT) approaches do not contain an evaluation of the dispersion contribution and hence are not suitable for modelling systems in which this contribution is dominant[108].

1.4.5 Short range interactions: Exchange-Repulsion

These interactions arise at short distances between two molecules, so that, an overlap between the molecular orbitals occurs, which generates strong repulsive forces. They arise from the Pauli principle, which states that two electrons cannot occupy the same region of space if they have the same spin. The electron exchange between the molecules gives weak attractive forces. The net repulsive force yields a modification of the charge density $\rho(r)$ of the system formed by the two atoms, which is decreased in the region between the two atoms and increased at the side of each atom remote from the other[94](figure 1.13).

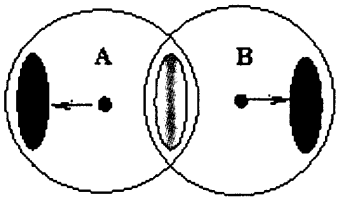


Fig. 1.13: Charge density of a system of two atoms, whose orbitals overlap. Grey colour indicates a decrease and black colour an increase of charge.

The overlap of the charge distributions also modifies the form of all the long range energy terms. In particular there is the *penetration energy*, which is the difference between the exact electrostatic energy calculated from the spatially distributed charge density as in [1.28] (using intermolecular perturbation theory method: IMPT[109,110]) and that derived from a converged multipolar expansion of the atomic charge density

DMA[111]. All these contributions decay approximately exponentially with increasing separation, resulting in a net repulsive energy term as:

$$U_{repulsion} = f(\boldsymbol{R}) \cdot e^{-\alpha \boldsymbol{R}} \ , \qquad [1.29]$$

where α is a constant and $f(\boldsymbol{R})$ is a function which varies slowly with the distance between the atoms. This term can be evaluated using a supermolecule approach[112] (see 1.5.1) or perturbation theory calculations. All of these methods are mainly based on the overlap model, which is the simple assumption that, short-range terms like repulsive energies are proportional to the degree of overlap S_ρ of the two molecular charge distributions:

$$U_{rep} = KS_\rho = K \int \rho^A(\boldsymbol{r}) \rho^B(\boldsymbol{r}) d\boldsymbol{r} \quad . \qquad [1.30]$$

The calculated data can be fitted by a model potential with the repulsive term of the form [1.29].

The repulsive part [1.29] of the intermolecular potential can be also fitted to empirical data, but the number of parameters to fit to a large number of experimental data is very limited, because the fitting only reflects the total forces acting on the rigid molecules at orientations sampled from the experimental crystal structure. The major problem is the blurring of the subdivision into contributions of the different atomic types implicit in the fitting (see section 1.5.2). In this case the overlap model can be helpful[111]. For high polarized, electronegative species (like chlorine[111]: Cl), with a non spherical charge distribution, a proper description of the repulsive contribution is given by the use of the anisotropic repulsive term, with the functional form depending on the orientation of the non spherical features of the charge distribution. The overlap method can estimate this orientational dependence, which should, where appropriated, be included as anisotropic repulsive terms in the intermolecular potential. Transferable anisotropic potentials were developed for many systems containing Cl···Cl interactions, such as *chlorobenzene* crystals[113] and they improved the ranking of the polymorphs of *p-dichlorobenzene*[113], for which the isotropic atom-atom potential cannot give a good crystal structure prediction[114].

1.4.6 Short range interactions: Charge Transfer & Damping of the Long Range Interactions

Another effect arising from the overlap of the atomic orbitals is the *charge transfer* from occupied orbitals of one molecule to unoccupied orbitals of the other. This term can be evaluated from the second order Rayleigh-Schroedinger perturbation theory[115] and is very

sensitive to the basis set chosen in *ab initio* calculations of the charge distribution of the molecule. It has an approximate exponential decay and it is extremely non-additive. Thus it is not usually explicitly included in potential models and is assumed to be partially absorbed in the exchange repulsive term. If large, this term would correspond to covalent bonding, which must be described in another way.

Molecular overlap also modifies the electrostatic, induction and dispersive interactions, which at short range cannot be described anymore solely as dependent on R^{-n}. This term in fact contains a singularity for $R \rightarrow 0$, which can be overcome by *damping*, which can be significant for van der Waals complexes. Different forms of the damping functions for long range interactions have been discussed[116].

1.5 Potential Models

Interatomic potential models describe the interactions, discussed above, in term of parameterised analytical functions. The parameters may be derived empirically, but are increasingly obtained by fitting the *ab initio* energies.

1.5.1 Super-molecular ab initio method

The *ab initio* methods solve Schroedinger's equation, without the use of empirical parameters. In applications to the potential energy, we use the "*super-molecule*" method[112]. Considering the simple case of a dimer, the method consists of calculating the interaction energy between two molecules A and B, from the energy difference between a system with the two interacting molecules, and the two isolated molecular systems:

$$U = U_{AB} - U_A - U_B \quad . \qquad [1.31]$$

The *ab initio* calculation uses the Born – Oppenheimer approximation, which considers the motion of the electrons in a field of fixed nuclei and is based on the Self-Consistent-Field (SCF) iterative procedure, which is the core of the Hartree-Fock method[89]. One big intrinsic problem in the pure Hartree-Fock method is that it completely neglects electron correlation, the main effect of which is to modify the electronic charge distribution around and between the molecules. It tends to reduce the charge separation, so that the atoms and bond charges are in effect smaller. In many cases, it is necessary to take in account these correlation effects, which can be done in different ways, using the MP2 correction[117] for the internal electrons correlation in the molecule and a Hartree-Fock-dispersion model (HFD)[118], in which an empirical expression for the dispersion is added

to an *ab initio* Hartree-Fock calculation, in which case the definition of the method as *ab initio* is debatable.

Secondly, the intermolecular interaction energy is usually very small (~10 kJ/mol.), much weaker than the energy involved in the chemical bonds and it requires careful consideration. By using the *variational principle*, which states that the energy of an approximate wave function cannot be lower than the lowest eigenvalue of the hamiltonian of the system, the SCF method guarantees the energy cannot be lower than the energy corresponding to the best wave function for the system. In the supermolecule method the variational principle does not give the same assurance, not giving us the confidence that the inevitable errors in the energy U_{AB}, and those in the energies $U_A + U_B$, will cancel out. Since the error in an *ab initio* calculation is of the order of 1eV per electron pair, for a molecular system we can reach errors (for U_{AB} and U_A+U_B) of the order of 10ev~1000kJ/mol, which is much larger than the order of the intermolecular interaction energy. A particular source of non-compensating errors is the *Basis Set Superposition Error* (BSSE), which arises when there is appreciable overlap between the two molecules, at short distances, in which case a part of the basis set for the molecule B, is used in the molecule A, to improve the description of the A's wave-function and *viceversa* in the case of the molecule B. The effect of the BSSE depends on the size of the basis set used in the calculation. The larger is the basis set, the lesser will be the BSSE[108]. An extensive study of the effect of the BSSE in different dimers is reported in a study of Rappe' & Bernstein[108].

Finally, a further complicating feature of the super-molecule model is due to the Born-Oppenheimer approximation, which is necessary to define an intermolecular potential. In this approximation the *ab initio* SCF calculation provides the final intermolecular interaction energy U for a certain fixed orientation of the molecules A and B. It is thus necessary to repeat the calculation of U for different relative orientations of the molecules A and B. The result is to define a potential energy surface (PES) which describes the intermolecular interaction energy as function of the spatial coordinates $U(R_i, \Omega_i)$. Such studies are computationally expensive even for dimers, and investigation of an entire crystal structure is often prohibitive. Finally, neither HF, nor Density Functional Theory[119] (DFT) consider correlations effects due to dispersion. At the present *ab initio* calculations at the MP2 level are used in calculating the interaction energy in dimers or trimers[108]. Usually calculations on dimers or trimers are performed at the MP2 level to include dispersion interaction, which are very sensitive to the basis set in including the

electronic correlations[120], but ongoing studies are trying to introduce a dispersive contribution in DFT methods[121,122].

1.5.2 Model Intermolecular Potentials

A potential U, containing solely the intermolecular interactions, is satisfactory for rigid molecular models, which treat the molecule as a rigid body (*rigid body approximation*). For high flexible molecules, containing high torsional and rotational degrees of freedom this model is limited and insufficient. In this case a potential containing an intramolecular contribution is necessary. Historically intra- and inter- molecular potentials were developed separately and care must be taken in assuring a full compatibility between the two parts[88].

Usually, in most force fields (inter + intra atom-atom potentials) the potential energy is partitioned in atom-atom contributions as[88]:

$$U_{tot} = \sum \frac{1}{2} k_{ij} \left(q - q^0\right)^2 + \sum \left[A_{ij} e^{-B_{ij} R_{ij}} + \frac{C_{ij}}{R_{ij}^{\ 6}} + \frac{F_{ij}}{R_{ij}} \right] . \qquad [1.32]$$

In the general expression [1.32] for a force field, the first, harmonic term corresponds to the intra-molecular part, with q denoting internal molecular vibrational coordinate (stretching and angle term), excluding the contribution from the torsions, while the second, anharmonic term indicates the inter-molecular contribution, with the first two terms representing respectively the repulsive and dispersive contribution, and the last term the electrostatic contribution, when only atomic point charges are used. A, B, C, D, F are fitted parameters and R_{ij} the distance between atom species belonging to different molecules. The first part is harmonic, whilst the second part is shallower and anharmonic. Different force fields have been developed for crystal structures of simple organic molecules and for proteins, including DREIDING[123], AMBER[124], GROMOS[125], CVFF[126], CFF[127], CHARMm[128], MM4[129]. Many of these force fields were developed for macromolecules or liquids, but they show certain transferability. A force-field for the study of flexible molecules like glycerol, has been entirely developed from *ab initio* calculations of all the different contributions[130], with good success, showing how crystal packing for flexible molecules is a balance between intermolecular and conformational energy, both of which require high accuracy in the model. In particular the AMBER force field has been used in a MD study of phase transition in cyclopentane, which is discussed in detail in chapter VII.

The intramolecular part is given by three components: the term expressing the bond stretching, which has the following expression:

$$\sum_{bonds} K_r (r - r_{eq})^2 \quad , \qquad [1.33]$$

where K_r is the spring constant and r_{eq} is the equilibrium bond length; the expression for the bond angle term in the harmonic representation is:

$$\sum_{bonds} K_\theta (\theta - \theta_{eq})^2 \quad , \qquad [1.34]$$

with K_θ and θ_{eq} being respectively the bending spring constant and the equilibrium bond angle; finally the expression for the torsional angle term is:

$$\sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] \quad , \qquad [1.35]$$

where V_n represents the energy barrier for the rotation, n the number of maxima (or minima) in one full rotation and γ is the angular offset. All of the values of the constants in the expressions for the bond length, bond angles and torsional angles are obtained by fitting to experimental or *ab initio* data. The electrostatic interactions for non-bonded interactions separated by exactly three bonds (e.g."C1-C4" interaction) are occasionally reduced by the application of a scale factor of 1/1.2.

The inter-molecular component involves the van der Waals interactions and the electrostatic terms. In particular concerning the former, a simple way to describe the repulsive-dispersive interactions is the Lennard-Jones (L-J) potential[131,132] with the form:

$$\sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] \quad , \qquad [1.36]$$

where R_{ij} represents the distance between two atoms belonging to different molecules and the sum is over all the pairs of atoms belonging to different molecules.

For rigid molecules many intermolecular potentials have been derived following the atom-atom potential method, proposed for the first time by Pertsin and Kitaigorodoskii[133], and they treat the van der Waals contribution including dispersion and exchange-repulsion terms. A typical sophisticated form among these potentials is the *6-exp* function, also known as Buckingham potential, with the analytical expression:

$$U = \sum_{ik} A_{\iota\kappa} e^{-B_{\iota\kappa} R_{ik}} - \frac{C_{\iota\kappa}}{R_{ik}^{6}} \quad , \qquad [1.37]$$

where atom i in one molecule and atom k in the other molecule are of types ι and κ respectively. The parameters $A_{\iota\kappa}$, $B_{\iota\kappa}$, $C_{\iota\kappa}$ are required to describe the interaction between

each pair of atoms, and are usually derived by fitting to a set of crystal structures and heats of sublimations for different compounds, or to *ab initio* calculated data.

Since a large number of observables would be necessary to fit parameters for all interactions, it is usual to calculate the heteroatom parameters using the following combination rules:

$$A_{\kappa\iota} = \sqrt{A_{\kappa\kappa} \cdot A_{\iota\iota}}, \quad B_{\kappa\iota} = \sqrt{B_{\kappa\kappa} \cdot B_{\iota\iota}}, \quad C_{\kappa\iota} = \left(\frac{C_{\kappa\kappa} + C_{\iota\iota}}{2} \right) \quad . \qquad [1.38]$$

An alternative expression for the van der Waals contribution is a function of the minimum energy separation $R_{\iota\kappa}^{0}$, well depth $\varepsilon_{\iota\kappa}$ and steepness parameter $\lambda_{\iota\kappa}$ of each individual atom-atom interaction.

$$U(R) = \sum_{ik} \frac{\varepsilon_{\iota\kappa}}{\left(1 - \frac{6}{\lambda_{\iota\kappa}}\right)} \cdot \left[\frac{6}{\lambda_{\iota\kappa}} e^{\alpha\left(1 - \frac{R_{ik}}{R_{\iota\kappa}^{0}}\right)} - \left(\frac{R_{ik}}{R_{\iota\kappa}^{0}} \right)^{6} \right] \quad . \qquad [1.39]$$

The trend of this potential $U(R)$ as function of R_{ik} is shown in figure 1.14:

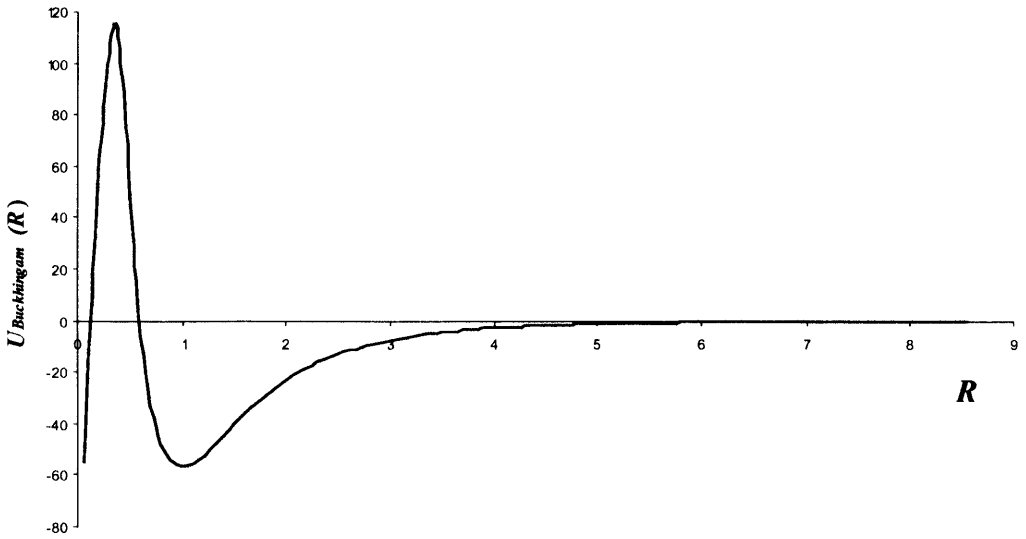


Fig. 1.14: *6-exp* atomic potential as function of the distance *R* between the atoms.

It is important to note the unphysical maximum in the *6-exp* potential, for very short distances. Nevertheless the system does not in practice sample these short distances, due to the very strong repulsive wall in the intermolecular energy.

Empirical fitting of an atom-atom potential is influenced by the different approximations in the calculation and the experimental errors in the measurements of the properties.

These factors affect the final potential, which is never exact in describing the real nature of the system. Filippini and Gavezzotti[134] have derived parameters for C, H, N, O, S and Cl by empirical fitting the *6-exp* form of interatomic potential without combining rules to a database of 217 crystal structures and heats of sublimation. The database includes hydrocarbons, oxahydrocarbons, azahydrocarbons, chlorohydrocarbons, sulfohydrocarbons, nitro compounds, sulfons and sulfoxides, but importantly not hydrogen bonded compounds. The method was extended only later to the latter[135]. These potentials reproduce adequately the experimental heats of sublimation for a large number of compounds; the calculated lattice energy is generally within ~10% of the experimental value. However the structures are predicted less well, with the cell edges frequently in error by as much as 5%. The limitation of this potential model is the lack of the explicit electrostatic contribution, which has been partially absorbed in the empirical parameters. The lack of explicit electrostatic contribution is an advantage in terms of computational cost, but it is not theoretically rigorous and it does not provide good transferability to other structures. The electrostatic contribution can become very important in particular strong directional interactions, which do not allow the molecular shape to play the major role in the determining the final crystal packing. Williams[136] extended his preliminary work with Starr[137] by examining the anisotropy in the electrostatic contribution and developing a method, which introduces additional *non*-nuclear sites until the ESP around the *ab initio* charge distribution is reproduced with acceptable accuracy. He improved his model from the study of azabenzenes[138] by including a variety of electrostatic models and always used chemical intuition to restrict the placing of the non-nuclear interaction sites. The final improvement was reached in 1999 and 2001 with a set of intermolecular potentials derived for hydrocarbons[139], oxohydrocarbons[140] and azahydrocarbons[141] in conjunction with an electrostatic point charge model fitted to the SCF 6-31G** *ab initio* derived potential, including various lone pairs . The potential, sometimes also denoted as W99, has a very good transferability and accuracy, by reproducing the crystal structure of carbohydrates, peptides and nucleosides with errors that are usually less than 2% in the lattice parameters. A key feature is the definition of atomic types by their intermolecular bonding, which distinguishes 2, 3, 4 coordinated carbons, carbonyl and 2 coordinated oxygens and nitrogens with 2, 1 or no bonded hydrogen atoms. A second important feature is the explicit repositioning of the hydrogen interaction sites at 0.1 Å from the nuclear site, in order to represent the displacement of the hydrogen charge distribution into the bond.

Another improvement is achieved with the FIT+0.9DMA potential derived by Coombes et al[142] for polar molecules, which comes from the Williams potential with an extension to consider different hydrogens (polar and non polar) and with an improvement of the electrostatic term. The latter contribution has been treated, by using atomic multipoles derived from the SCF 631G* wave function. The multipole expansion is up to the order of R^{-5}, using the Ewald summation (see section 3.3) for the charge-charge, charge-dipole, dipole-dipole contributions to the lattice energy and a direct summation on all the atoms in the molecules within a cutoff of 20Å. A scale factor of 0.9 in the electrostatic contribution, originating from a DMA approach, was used in order to compensate for the overestimated molecular dipoles, because of the lack of electronic correlation in the SCF method. The repulsive-dispersion contribution is derived by fitting of the parameters to empirical data for ~ 20 compounds. A distinction is made between the hydrogen atoms bonded to carbon, H_C, and those bonded to more electronegative elements (as nitrogen), H_P, in order to take into account the major difference in the electron charge distribution associated with the two types of hydrogen atoms. The empirical parameters for C, N and H_C were taken from the potential developed by Williams and Cox[136].

The FIT+0.9DMA potential was tested on 40 crystal structures of polar molecules and reproduces well 37 of these compounds, for which lattice parameters are reproduce within an error less than 2% with a poor representation only for benzene, *s*-tetrazine and indazole[142], which show particular sensitivity to the presence of anisotropy in the repulsive part of the intermolecular potential. The use of atomic point charges in treating the electrostatic contribution in the case of polar molecules gives worse results than the DMA. An important point to underline is the rigidity of all these compounds. In fact, the DMA method implies a rigid body approximation in treating the molecule, such that for highly flexible molecules, a conformational analysis of all the possible stable conformations and a separate calculation of the multipole expanded electrostatics for each of the conformations, is necessary. This procedure is computationally expensive and adds to the intrinsic computational expense of lattice energy minimisation with the use of multipoles. Indeed crystal structure prediction with the use of rigid body approximation + DMA model for flexible molecules can be very expensive in terms of computer time. Attempts were made to couple a multipole based electrostatic model with a force-field, containing intramolecular contributions for flexible molecules, permitting the relaxation of the molecular geometry during the optimisation[143]. Other techniques have been applied, with encouraging results, although being very demanding in terms of time and

computational cost[144,145]. For molecules without high flexibility an improvement in the description of the system is given by the use of the multipoles instead of simple point charges, but in general the flexibility is still a source of problems and inaccuracy in reproducing the experimental structures of such compounds.

1.6 Search Methods

Dzyabchenko[146] and Gavezzotti[147] undertook pioneering work on the prediction of crystal structures. The first, starting from 200 initial guesses in four space groups proposed seventeen structures for benzene, with encouraging results[146]. Gavezzotti[147] performed a search of the crystal structures of organic crystals by building clusters with the most common symmetry elements observed in the organic solid state: inverse centre, screw axis, glide plane. The search was guided by a statistical analysis of the packing motifs in the crystal structures of known hydrocarbons.

Nowadays there are different approaches adopted by different groups in attempting to predict crystal structures of molecular crystals. The differences concern many aspects of the method, from the different algorithms used in sampling the space of all the possible structures, to the algorithm used in the energy minimisation and the intermolecular potential used in modelling the system. Here we follow a classification proposed by Verwer and Leusen[148].

1.6.1 Formation of Molecular Clusters

Clusters of 10-50 molecules are formed and their energy is minimised. This method simulates the crystallisation process without assuming periodicity in the structures. Williams used this method, in predicting crystal structures of benzene[149], by performing energy minimisation of clusters up to 15 molecules. The weak point of these methods is the cluster dimension, which should be extended to a much bigger number of molecules to give better results.

1.6.2 Formation of clusters through symmetry elements

The basic assumption of this method is that particularly strong, directional intermolecular interactions, like hydrogen bonds or π interactions, can be recognised before the formation of the entire crystals, being already present in the smaller units, like dimers, or trimers, of the crystal structure. Initially, small clusters of 1 -10 molecules are formed by applying specific chosen symmetry operators. The most promising clusters are replicated

through translational operators to form crystals. The bulk environment is simulated, by imposing periodicity. Once identified, all these smaller units are symmetry translated, in order to reproduce the bulk crystal. Finally the complete structure is energy minimised. Examples of this methodology are the programs PROMET[150] and FlexCryst[151-153], which have respectively the limit of having a very low level of accuracy in modelling the electrostatic contribution (PROMET) and of relying upon the statistical frequency in the occurrence of specific motifs (FlexCryst).

1.6.3 Formation of trial structures

The key point of this group of methods is that periodicity is assumed at all stages of the different procedures to search for crystal structures.

The methods differ a lot in the initial starting points, in the methodologies for sampling the lattice energy minima space, in the algorithm used for the final lattice energy minimisation and in the number of variables involved in the minimisation. One method (*MDCP*)[154-156] for example uses molecular dynamics (MD) simulations in an anisotropic constant pressure ensemble, to allow a small number of molecules find the most stable configurations, by moving in a flexible unit cell, under periodic boundary conditions. The starting point consists of the placement of a rigid "seed" molecule at the centre of a "roomy" cubic cell followed by the application of different symmetry elements of the space groups to generate the other molecules in the unit cell. The method does not have an initial *a priori* assumption as to the space group and the symmetry of the crystals. All the final structures generated are energy minimised. This method can include the entropic effects based on the harmonic approximation. It achieved a distinct success in the crystal structure prediction of CO_2 and was successful in predicting the crystal structures of benzene, but not for pyrimidine. The risk of having the system trapped in local minima is always present in this method.

Other methods are based on a grid search, in which a systematic search is performed by placing molecules on a specific grid of points, with different symmetries, corresponding to the most common space groups in the organic solid state. One of these program UPACK[157] is able to handle intramolecular interactions, allowing bond extensions, bond angles, torsion angles an improper torsion angles to be part of the potential energy function. In the initial random search, a simple potential, consisting of solely repulsive terms, is used. The best candidate structures are lattice energy minimised and are considered only if the decrease in the lattice energy reaches a certain threshold. Up to this

stage, the molecules are treated as rigid. The final minimisation of the chosen structures is performed with a better potential, which includes hydrogen bond lengths and the torsions of the groups as minimisation variables. In this final minimisation the electrostatic contribution is usually also taken into account in the force field and is modelled with a distributed multipole expansion. Moreover, the method allows us to relax the space group symmetry by expanding P1.

An extended version of the program permits a starting configuration with more than one molecule in the asymmetric unit cell[158] and better force fields to treat flexible molecules[159], including a correction for electrostatics and the polarization[160]. In the case of flexible molecules, all internal degrees of freedom are considered as variables and the local minimisation is interrupted whenever the variation of the torsion is bigger than a certain amount[161]. The multipole moments are recalculated and the minimisation is then restarted. Nevertheless this procedure is computationally expensive. The application to crystal structures of methane, ethane, 1,4 dioxane and propane gave encouraging results[162].

Two other similar methods, based on a systematic grid search of structures starting with initial definition of the space groups, are ICE9[163] and MOLPAK[164]. A molecule is placed in the centre of an asymmetric unit and the construction of a coordination sphere of other molecules with different space group symmetry elements is performed by varying the orientation of this molecule in all the space of the possible orientations, with a systematic increment of certain degrees. Having determined all the candidate structures, a lattice energy minimisation is performed, with respect to the following variables: rotation, translation of the asymmetric unit, the three lattice vectors and the three angles. Both of the programs can treat only one rigid molecule in the asymmetric unit cell ($Z' = 1$). The main difference between the two programs is the use of a central multipole expansion in ICE9[163], instead of a distributed one and the use of a L-J potential[131] instead of a *6-exp* in treating the repulsive-dispersive contribution. The crystal structure prediction for ICE9 is successful for a certain number of aromatic hydrocarbons, but failures are observed for molecules with irregular shape and containing a certain internal flexibility such as bicyclohexilidene[163]. The main obstacle in finding a global minimum is given by the energy barrier of molecular rotations between local minima. The MOLPAK[164] method (a more complete description of which is given in chapter 3, section 3.1), combined with a careful lattice energy minimisation of the generated structures with DMAREL[165] (a description of this program is given in chapter 3, section 3.2), has given good results in

searches of many compounds as well as in the prediction of two of the three polymorphs of 2-amino-5-nytropyrimidine[166]. The hydrogen bond motif observed in the two experimental forms is the same for the global minimum structure, while the third form was not reproduced, due to the flexibility of the molecule, which had a different experimental conformation from the gas phase conformation, considered in the prediction. The method was also able to reproduce different structures of small carboxylic acids[167], azauracil, allopurinol[168] and to predict the most stable structure of paracetamol as the global minimum, with the second ranked structure very close to the less stable orthorhombic polymorph[169].

Other methods are based on a stochastic search of the lowest energy minima, by assuming constant space group symmetry. This procedure implies that the orientations of the independent asymmetric cell and the lattice parameters are sufficient to generate all atomic coordinates. One of these methods, called *Polymorph Predictor*[170,171], is based on a Monte Carlo method, and the probe of the space of the possible structures is performed, by varying the orientations of the independent asymmetric units and the lattice parameters. The energies of the generated structures are computed and a criterion based on the Metropolis algorithm[172] is adopted for the acceptance or rejection of a generated structure. Given its stochastic nature the method requires the search to be repeated until no new structures are generated. The final stage of the search consists of a lattice energy minimisation of the generated structures, by using an Ewald summation[173] to treat the electrostatic contribution and second derivatives. The space group symmetry is retained at all stages. This method can also consider more than one independent molecule in the asymmetric unit cell, but in the case of flexible molecules each low energy conformation has to be searched separately, given the rigid body approximation. Payne et al.[174] used this method to predict the crystal structure of primidone and progesterone, with partial success, but were unable to locate the structure as global minimum in lattice energy. This method has shown difficulties in probing the search space of the possible structures in a complete way, especially in the cases of $Z' > 1$ systems and of flexible molecules.

To improve the sampling of the search space within methods based on a stochastic search of the low energy structures, programs were developed to use a quasi-Monte Carlo approach with a low-discrepancy sequences method[175]. One of these programs gave a successful prediction of the pentacene compound[176], with the two experimental polymophs corresponding to the two lowest energy minima in the search. Another[177] based on a quantum-Monte Carlo search, improves the description of the electrostatics

(usually treated with atomic point charges in these stochastic methods) by using satellite charges in different points of the molecule, which are determined by optimisation of the deviation of the deformed electrostatic potential from that calculated *ab initio* from the wave function[178]. This improvement gave very good results as demonstrated by the successful prediction of triethylenediamine[177]. The program also allows for the relaxation of the space group symmetry in the final energy optimisation of the structures produced by the search and can deal with more than one molecule in the asymmetric unit cell. Encouraging results were obtained in the prediction of various compounds, some of them with $Z' > 1$, although often the experimental structure is not found as a global minimum in the search, as in the case of allopurinol[177].

Other different approaches are followed to predict the crystal structures of organic molecules. The program CRYSCA[179] is based on the energy minimisations of structures obtained by random packing, allowing intermolecular degrees of freedom. It uses a potential, which contains a *6-exp* term plus an additional point charge Coulombic term. The minimisation is performed with respect to molecular positions, orientations and intramolecular degrees of freedoms.

Another method is an extension of the Monte Carlo method, known as Conformational Family Monte Carlo[180] (CFMC). The central point in this method is the structure family database, which is an ensemble of structures grouped into different families. In contrast with the simulated annealing Monte Carlo, this method does not use a single structure in the search of possible local minima, but an entire group of structure (a family) and only moves between families are accepted or rejected, using a Metropolis algorithm criterion. Finally a lattice energy minimisation is performed to bias the search toward the regions with lowest energy families. AMBER[132] and W99[140] potentials were used in the energy minimisation procedure, treating the electrostatics using only atomic point charges. The method performs quite successfully in the case of rigid molecules, although a fairly rigid molecule such as imidazole does not seem to reproduce well the lattice parameters[181]. Furthermore the method is not really successful in the case of highly flexible molecules, with a large number of torsional parameters, due to the fact that the search does not cover properly all the conformational space.

The MPA/MPG[182] program developed by D.E. Williams is able to find the lattice energy minima for a certain number of molecules in a unit cell without any assumption as to the space group symmetry and using many different algorithms of lattice energy minimisations. Programs based on diffusion equation methods[183,184] are also used in

crystal structure prediction. Essentially, all these methods attempt to transform the multidimensional lattice energy surface of a system, removing all the insignificant minima, and trying to merge all shallow minima into the deepest basins, which would finally represent all the possible structures for the compound. These can be tracked back by applying a reversing procedure, which gradually removes the deformation in the lattice energy surface. All the molecules are treated as rigid bodies and can move freely in the unit cell. No space group constraint is imposed.

Finally a modified genetic algorithm method[185] was applied for the prediction of crystal structures, with the latter coded in real-valued vectors, which are a part of the *genome* of the structure. The generation of all the possible structures is performed, by using the *genome*: initial crystal blocks are constructed for each individual in the population. Their potential energy is evaluated and the structures are relaxed with a local minimisation. Individuals with positive energy are eliminated from the population. Operations of mating (crossover), mutation and selection are used to evolve one generation into the next and a convergence criterion is defined by the relative difference in lattice energy between successive cycles. The potential energy used is a non bonded atom-atom potential in the *6-exp* form plus a partial charge point charge model for electrostatics, treated with the Ewald summation. This method predicted successfully the structures of benzene, naphthalene and anthracene[185], but has shown some limitations in treating flexible molecules.

1.7 The State of Art in CSP

All the methods, previously discussed, of CSP of simple organic molecules have reported some successes, and show a good ability in reproducing the structure of many important compounds, including those of industrial importance, such as aspirin[186], paracetamol[169](pharmaceutical) and quinacridone[187] (pigment). Nevertheless all of them show some failures, so that none can be considered "The" method to predict crystal structures. In the case of the molecules whose structures were predicted, the term "success" for the prediction has to be carefully interpreted. A truly successful method of CSP should predict as lowest energy structures, all the effectively experimental observable (and not observed) structures (under the assumption of thermodynamic control, as discussed in section 1.2.1), reflecting their thermodynamic stability and giving the conditions for which one structure can be observed rather than another. This issue is

far from being obtained by the current methods of CSP. A first difficult point is that many of the predicted structures for molecules like acetic acid[188] or aspirin[186], were not ranked as those with lowest energy, and from the point of view of a pure prediction this result would not be a success. Still the prediction of the structure as one of the most stable is a very encouraging result and it can be considered a success[63]. The more fundamental question concentrates on establishing if the other lowest energy predicted minima are effective possible polymorphic forms and, in the case of a more stable predicted structure, if the observed form is metastable. This problem is very difficult to solve, essentially because, as widely recognised[62,63,187,189] a typical search of crystal structures of a certain organic compound produces a larger number of local minima in the lattice energy in a very small range of energy ($\leq$ 5kJmol^{-1}) than the number of the known polymorphs. As mentioned before, even with the perfect *ab initio*, fully transferable intermolecular potential we would not be able to decrease the number of local minima produced in a CSP, by only considering the thermodynamic aspect of the problem[157]. At present, it is widely recognised that there is a good capability for the current methods in predicting crystal structures of rigid molecules with $Z'=1$. On the other hand, flexible molecules represent still a big challenge in CSP. In fact, if the stability of a particular molecular conformation reduces the number of possible structures that can be crystallised for flexible molecules, on the other hand, from the point of view of theoretical crystal structure prediction, an even bigger number of possible local minima is likely to occur, compared with the case of a rigid molecule[190]. Even in the case of limited molecular flexibility, the CSP becomes very complicated, due to the presence of a wide variety of local minima structures, as observed for barbituric acid[191], with a variety of hydrogen bonded motifs and packing energies. Often the molecular conformation in the crystal structure is slightly less stable than that in the gas phase, but this loss in intramolecular energy can be easily balanced by a reduction in the intermolecular energy of the crystal. The total energy of the crystal is in this case given by the sum of the two contributions:

$$U_{tot} = U_{latt} + \Delta U_{intra} \quad . \qquad [1.40]$$

Small differences between the different molecular conformations can produce important changes in the relative final energies U_{tot} of two possible structures (~5/8kJmol^{-1}), which determine the range in which polymorphism occurs[189,191]. The large influence on the lattice energy, given by small changes in the molecular geometry is well illustrated by the case of uric acid[192], whose experimental structure was located as a local, less dense structure 5kJmol^{-1} above the predicted global minimum in a CSP with

MOLPAK/DMAREL[164,165], by using the *ab initio* gas phase optimised molecular conformation. A structure with a different hydrogen bond motif was predicted as the global minimum in the search. Nevertheless a CSP with the same program, by using the experimental molecular conformation predicted the observed structure as the global minimum, although the different hydrogen bonded motifs are still predicted as local minima in a range of 5kJmol^{-1}. These alternative structures may be possible different polymorphs of uric acid, whose crystallisation is prevent by the very poor solubility of this molecule.

Another critical issue in CSP is the possibility of crystal structures with more than one molecule in the unit cell. In the recent blind test of 2004, the compound XI, initially thought to be easily predictable, showed this behaviour and its structure was not predicted by any of the participants. In the case of more than one molecule in the unit cell the number of minima produced in the search increases and the problem of sorting structures and elimination of similar structures becomes even more critical[193], as observed in the CSP of the 2004 blind test molecule XI (azetidine), which is fully discussed in chapter 5. Generally the major problem in the prediction of crystal structure of organic molecules is in the occurrence of a very large number of local minima within a small energy range (~ 5kJmol^{-1}). Some possible outcomes of the lattice energy landscape from a CSP are shown in figure 1.15:

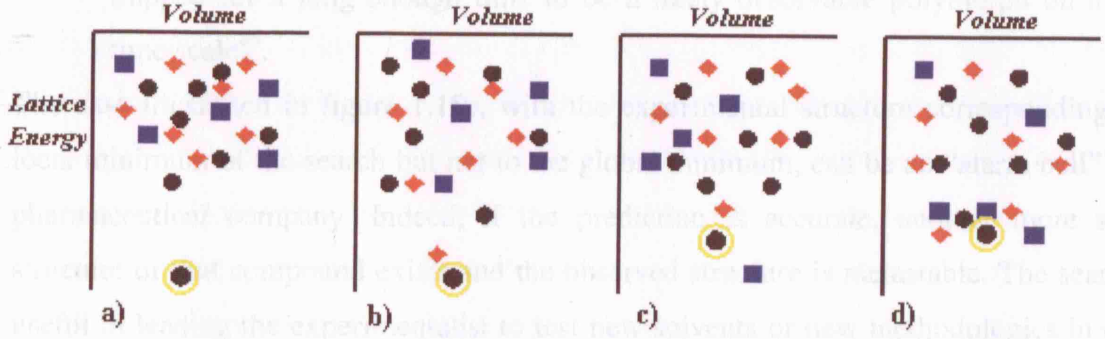


Fig.1.15: Schematic representation of the different possible lattice energy landscapes from a computational CSP. Each point corresponds to a specific crystal structure. The different colours and shapes of the energy minima correspond to different space groups. The yellow circle represents the experimental structure.

If the predicted global minimum corresponds to the experimental structure and there is a big energy gap between this structure and the other local minima in the energy landscape,

then we can say that the prediction was unequivocally successful and there is no polymorphism in the system (figure 1.15a). If the experimental structure corresponds to the global minimum of the search but only a small gap separates this one from other possible minima (figure 1.15b) it is opportune to make careful considerations about:

- The similarities and difference in the motifs and packing of the other minima with respect the global minimum. Many different programs have been developed to compare different structures in order to establish similarities and differences among each others[194,195]. Also the use of calculated X-ray powder patterns[196,197] and crystal engineering can have a useful application at this stage in the prediction and in further investigations (see paragraph 3.6).
- The possibility of re-ranking of the energies at room temperature, by including thermal effects, in order to establish the nature of the polymorphic system: enantiotropism or monotropism[85].
- The answer to the questions: are the other structures able to be crystallised, for specific reasons? Are these minima characterised by a specific structure, which considerably differs from that of the global minimum (i.e. different hydrogen bonds motif), such that a big energy barrier separates the two different structures? In the latter case, even if thermodynamically metastable, once kinetically obtained, the structure corresponding to a certain local minimum would be trapped for a long enough time to be a likely observable polymorph on a long time scale[85].

The case illustrated in figure 1.15c, with the experimental structure corresponding to a local minimum of the search but not to the global minimum, can be an "alarm bell" for a pharmaceutical company. Indeed, if the prediction is accurate, another more stable structure of that compound exists and the observed structure is metastable. The search is useful in leading the experimentalist to test new solvents or new methodologies in order to search another possible structure, which can be found also considering the "supra-molecular synthon" of the predicted structure.

Finally the case, illustrated in figure 1.15d, in which there is not a clearly distinct global minimum, but many different local minima that are very close in energy to the global minimum, is the most difficult scenario in CSP. The prediction is not trivial, but the results can still be useful in developing strategies to find new polymorphs and characterise them with X-ray powder diffraction analysis.

The first essential requirement for a useful CSP is the quality of the energetic ranking of the predicted structures. Indeed, two observed polymorphs can often have very small differences in lattice energy (~ 2 kJ mol^{-1}). The small energy differences between two polymorphs require a very accurate ranking of the structures. This is possible only with an accurate force field for modelling the intermolecular forces involved in the system with wide transferability[63]. This problem is far from being solved. The various *ab initio* methods applied to improve the important electrostatic contribution, like DMA[97], satellite charges[178] and pixel methods[63,198,199], have given distinct improvements in CSP for structures dominated by strongly directional interactions, like hydrogen bonds. The pixel method also enables an *ab initio* and inexpensive treatment of polarisation, whose contribution, which is often neglected in CSP with empirical potentials, can play a very important role[63]. However an important contribution from the repulsive-dispersive may play a deciding role in many systems, for which anisotropy in the repulsive part can dramatically affect the result of the prediction. The *ab initio* model of this contribution is still far from being accurately reached[63].

Once the problem of the "ideal" force field is solved there will still remains a major problem: the large number of local minima in a very restricted range of energy landscape (cases shown in figure 1.15b, 1.15c, 1.15d). In fact the "ideal" force field can guarantee a very accurate ranking of the lattice energy local minima, but cannot decrease their number. This characteristic is very common and observed in the most of organic crystals and is indeed the main limiting factor in the computational CSP of such materials. Indeed by using only a purely thermodynamic approach, restricted to calculations of the lattice energies at 0K, we are neglecting very important factors competing in the appearance of a polymorph, like thermal and entropic effects at room temperature and kinetic factors. The former can sometimes be partially included in parameters fitted to empirical data at room temperature. A more appropriate way to include it is by calculating the entropic effect arising from all the phonon vibrations in the crystal, including the zero point energy. By multiplying this term by a certain temperature and adding the result to the lattice energy, we can get an estimate of the free energy of a particular structure at that temperature. The calculation of the vibrational frequencies of the crystal involves the calculation of the second derivatives of the energy with respect to the lattice constants and atomic positions of the system, in the quasi-harmonic approximation (section 3.8). This method does not take into account anisotropic effects in the vibrations of the solid a certain temperatures and is only an approximate evaluation of the free energy of the system.

At present, the problem of simulating the kinetics of crystal formation is still far from being solved, because of the lack of an accurate model for the nucleation and crystallisation of a crystal, including factors[63], like the nucleation rate and, solvent effect. Different attempts at a computational study of the pre-stages in nucleation, have been performed by using the molecular dynamics (MD) method[200,201], which is in principle ideal to give insights about the kinetic behaviour of the formation of a crystal; but there are still big limitations in time cost for MD and it is not possible to run simulations for time as long as the time scale of a nucleation and crystal growth, although very useful insights can be gained[44].

Another attempt to take into account, with a very crude approximation, the kinetic factor in the crystallisation of the different polymorphs is represented by the calculation of the volume growth rate of a crystal, in vacuo[169,202]. These kinds of simulations are based on the prediction of the morphology of the crystals in vacuo and the method is based on the attachment energy model. A more detailed description of the method is given in appendix 3.3, chapter 3. Although with many limitations, this method can be useful in reducing the number of local minima in the search, by removing structures, which show a particularly slow volume growth rate. Both of these methods, morphology calculation and MD are based on the use of intermolecular potentials; hence they are sensitive to the accuracy of the potential.

Finally, another useful tool in reducing the number of local minima in the energy landscape is given by the calculation of the elastic constants of the structure, using second derivatives of the lattice energy, in order to estimate the mechanical stability of a crystal[84,169]. Structures with particularly low values of even one of the lattice constants, indicate particular instability to specific mechanical deformations and hence are unlikely to crystallise, even if present in solution at the very initial stages of the crystallisation. A more detailed explanation of the method is treated in appendix 3.2, chapter 3

In conclusion, the CSP of polymorphs of organic molecular crystals is still far from being a solved problem. Going back to the direct question, which is the title of two papers[61,62] published at a distance of almost ten years one from each other, "*Are crystal structure predictable?*", the direct answer to this question at the current state of art in CSP can be: "*Sometimes...*". Indeed considerable progress was achieved in these years, with interesting successes and a considerable number of compounds for which CSP is possible with the present developed methodology[83]. Furthermore even in those cases of failure, CSP methods still have provided useful guidance to experimental investigation of

polymorphic systems. Good interaction between experimental and theoretical investigations has produced very useful results in some systems such as paracetamol[20], 5-fluorouracil[203] and progesterone[204] and in giving useful information and clues on the causes of disappearing polymorphic forms[205]. Moreover, increasing interaction of the complementary approaches of computational crystal structure prediction and crystal engineering can give very good results in predicting and understanding polymorphism for certain kind of systems, as well represented by the successful prediction of the crystal structure of molecule IV in the blind test 2001 (figure 1.18), by Sarma et al[206]. These results indicate how the interaction of the different complementary approaches will be essential in "realising" CSP of the organic solid state.

In this thesis, we will illustrate the complexity of computational CSP, considering the work performed on three different kinds of very simple molecules, which represent different cases in the landscape of the organic compounds previously discussed. The molecules are: imidazole, a highly rigid molecule with strongly polar groups, giving intermolecular hydrogen bonded interactions, azetidine (compound XI of the blind test 2004: see paragraph 5.1.1), which is a small slightly flexible molecule, containing a polar group, which can have important electrostatic intermolecular interactions and finally cyclopentane, a flexible molecule without any polar group, whose solid state is likely to be characterised by weak intermolecular interactions, with dominating dispersive contributions. Although these molecules are relatively simple, their CSP is far from trivial and raises many important issues and questions, which will be discussed in the respective chapters.

1.8 References

(1) Mc Crone, W. C. *Physics and Chemistry of the Organic Solid State*; Fox, D.
Labes, M.M.
Weissberger, A.: New York, 1965; Vol. II.

(2) Raiteri, P.; Martonak, R.; Parrinello, M. *Angew. Chem. Int. Edit.* **2005**, *44*, 3769.

(3) Garnier, P.; Gregoire, P.; Montmitonnet, P.; Delamare, F. *J. Mater. Sc.* **1988**, *23*, 3225.

(4) Zanotti, F.; Masiello, S.; Bader, S.; Guarneri, M.; Vjnovic, D. *Int. J. Cosmet.Sc.* **1998**, *20*, 217.

(5) Erk, P. *Curr. Op. Solid State Mater. Sc.* **2001**, *5*, 155.

(6) Fehr, C.; Dieudonne, P.; Primera, J.; Woignier, T.; Sauvajol, J. L.; Anglaret, E. *Eur. Phys. J.* **2003**, *12*, S13.

(7) Borka, L. *Pharm. Acta Helv.* **1991**, *66*, 16.

(8) Borka, L.; Haleblian, J. K. *Acta Pharm. Jugosl.* **1990**, *40*, 71.

(9) Caira, M. R. Polymorphism In *Encyclopedia of Supramolecular Chemistry*; Dekker, M., Ed., 2004; pp 1129.

(10) Wilding, M. C.; Wilson, M.; McMillan, P. *Chem. Soc. Rev.* **2006**, *35*, 964.

(11) Threlfall, T. L. *Analyst* **1995**, *120*, 2435.

(12) Amelincks, S. *Acta Cryst.* **1956**, *9*, 16.

(13) Dunitz, J. D. *Pure Appl. Chem.* **1991**, *63*, 177.

(14) Doyle, B.; Hukins, D. W. L.; Hulmes, D. J. S.; Miller, A.; Woodhead-Galloway, J. *J. Mol. Biol.* **1975**, *91*, 79.

(15) Jackson, A. P.; Maxwell, A.; Wigley, D. B. *J.Mol.Biol.* **1991**, *217*, 15.

(16) Chapman, D.; Urbino, J.; Keough, K. M. **1974**, *249*, 2512.

(17) Nesrullajev, A.; Salihoglu, S.; Yurtseven, H. *Int. J. Mod. Phys. B* **1998**, *12*, 213.

(18) Corradini, P.; Guerra, G. *Adv. Polymer Sci.* **1992**, *100*, 182.

(19) Bernstein, J. *Polymorphism in Molecular Crystals*, 2002.

(20) Bernstein, J. *ACA Trans.* **2004**, *39*, 14.

(21) Haleblian, J.; McCrone, W. *J. Pharm. Sc.* **1969**, *58*, 911.

(22) Allen, F. H.; Motherwell, W. D. S. *Acta Cryst. Sec. B* **2002**, *58*, 407.

(23) Byrn, S. C. "Diversity Amidst Similarity: A Unified Approach to Polymoprhs Solvates and Phase Changes"; 35th Erice's School of Crystallography, 2004, Erice, Sicily.

(24) Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. *Adv. Drug Del. Rev.* **2001**, *48*, 3.

(25) Singh, D.; Marshall, P. V.; Shields, L.; York, P. *J. Pharm. Sci.* **1998**, *87*, 655.

(26) Blagden, N.; Davey, R. J.; Lieberman, H. F.; Williams, L.; Payne, R. S.; Roberts, R. J.; Docherty, R. *J.Chem.Soc.,Faraday Trans.* **1998**, *94*, 1035.

(27) Sledz, M.; Janczak, J.; Kubiak, R. *J. Mol.Struc.* **2001**, *595*, 77.

(28) Nobeli, I.; Price, S. L. *J. Phys. Chem. A* **1999**, *103*, 6448.

(29) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.

(30) Beyer, T.; Price, S. L. *J. Phys. Chem. B* **2000**, *104*, 2647.

(31) Caira, M. R.; Zanol, M.; Peveri, T.; Gazzaniga, A.; Giorndano, F. *J. Pharm. Sc.* **1998**, *87*, 1608.

(32) Desiraju, G. R. *J.Chem.Soc.Perkin Trans. 2* **1983**, 1025.

(33) Ferraris, J.; Walatka, V.; Perlstei.Jh; Cowan, D. O. *J. Am. Chem. Soc.* **1973**, *95*, 948.

(34) Dromzee, Y.; Chiarelli, R.; Gambarelli, S.; Rassat, A. *Acta Cryst. Sec. C* **1996**, *52*, 474.
(35) Miller, J. S.; Epstein, A. J. *Chem. Commun.* **1998**, 1319.
(36) Miller, J. S.; Epstein, A. J. *Angew. Chem.-Int. Edit. Engl.* **1994**, *33*, 385.
(37) Epstein, A. J.; Miller, J. S. *Synth. Met.* **1996**, *80*, 231.
(38) Law, K. Y. *Chem. Rev.* **1993**, *93*, 449.
(39) Hall, S. R.; Kolinsky, P. V.; Jones, R.; Allen, S.; Gordon, P.; Bothwell, B.; Bloor, D.; Norman, P. A.; Hursthouse, M.; Karaulov, A.; Baldwin, J.; Goodyear, M.; Bishop, D. *J. Cryst. Growth* **1986**, *79*, 745.
(40) Begley, M. J.; Crombie, L.; Griffiths, G. L.; Jones, R. C. F.; **Rahmani, M.** *J. Chem. Soc.-Chem. Comm.* **1981**, 823.
(41) Reetz, M. T.; Hoger, S.; Harms, K. *Angew. Chem.-Int. Edit. Engl.* **1994**, *33*, 181.
(42) Bernstein, J.; Davey, R. J.; Henck, J. O. *Angew. Chem. Int. Edit.* **1999**, *38*, 3440.
(43) Gavezzotti, A. *Theoretical Aspects and Computer modeling of the Molecular Solid State*; John Wiley & Sons: Milan, 1997.
(44) Gavezzotti, A. *Cryst. Rev.* **1998**, *7*, 5.
(45) Gavezzotti, A. *Modelling Simul. Mater. Sci. Eng.* **2002**, *10*, R1.
(46) Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1995**, *117*, 12299.
(47) Burger, A.; Ramberger, R. *Mikrochim. Acta* **1979**, *2*, 273.
(48) Grunenberg, A.; Henck, J. O.; Siesler, H. W. *Int. J. Pharm.* **1996**, *129*, 147.
(49) Mullin, J. W. *Crystallization*, 4th ed.; Butterworth-Heinemann, Ltd.: Oxford, 2001.
(50) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601.
(51) Davey, R. J.; Garside, J. *From molecules to Crystallizers: An Introduction to Crystallization*, 2000.
(52) Dunitz, J. D.; Bernstein, J. *Acc. Chem. Res.* **1995**, *28*, 193.
(53) Bombicz, P.; Czugler, M.; Tellgren, R.; Kalman, A. *Angew. Chem.-Int. Edit.* **2003**, *42*, 1957.
(54) Read, J.; Smith, G. S. *J. Chem. Soc.* **1921**, *119*, 779.
(55) Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. *Org. Process Res. Dev.* **2000**, *4*, 413.
(56) Blagden, N.; Davey, R. J.; Rowe, R.; Roberts, R. *Int. J. Pharm.* **1998**, *172*, 169.
(57) Gu, C.-H.; Chatterjee, K.; Young Jr. , V.; Grant, D. J. W. *J. Cryst. Growth* **2002**, *235*, 471.
(58) Kitamura, M. *Cryst. Growth & Des.* **2004**, *4*, 1153.
(59) Maddox, J. *Nature* **1988**, *335*, 201.
(60) Erk, P.; Hengelsberg, H.; Haddow, M. F.; van Gelder, R. *CrystEngComm* **2004**, *6*, 474.
(61) Gavezzotti, A. *Acc.Chem. Res.* **1994**, *27*, 309.
(62) Dunitz, J. D. *Chem. Comm.* **2003**, 545.
(63) Gavezzotti, A. *CrystEngComm* **2002**, *4*, 343.
(64) Kitaigorodskii, A. I. *Organic Chemical Crystallography*New York: Consultant Bureau, 1961.
(65) Gavezzotti, A. *J. Am. Chem. Soc.* **1985**, *107*, 962.
(66) Kitaigorodskii, A. I. *Acta Cryst.* **1965**, *18*, 585.

(67) Kitaigorodskii, A. I. *Molecular Crystals and Molecules*New York, 1973.
(68) Motherwell, W. D. S. *Acta Cryst. Sec. B* **1997**, *53*, 726.
(69) Wilson, C. C. Space Groups and Crystal Packing Modes. In *Encyclopedia of Supramolecular Chemistry*; Dekker, M., Ed., 2004; pp 1337.
(70) Desiraju, G. R. *Crystal Engineering: The design of molecular solids*Amsterdam, 1989.
(71) Sarma, J. A. R. P.; Desiraju, G. R. *Cryst. Growth & Des.* **2002**, *2*, 93.
(72) Allen, F. H. *Acta Cryst. Sec. B* **2002**, *58*, 380.
(73) Coombes, D. S.; Nagi, G. K.; Price, S. L. *Chem. Phys. Lett.* **1997**, *265*, 532.
(74) Leiserowitz, L. *Acta Cryst. Sec. B* **1976**, *32*, 775.
(75) Leiserowitz, L.; Hagler, A. T. *Proc. R. Soc. Lond. A* **1983**, *388*, 133.
(76) Leiserowitz, L.; Tuval, M. *Acta Cryst. Sec. B* **1978**, *34*, 1230.
(77) Taylor, R.; Kennard, O. *J. Am. Chem. Soc.* **1982**, *104*, 5063.
(78) Berkovitch-Yellin, Z.; Leiserowitz, L. *Acta Cryst. Sec. B* **1984**, *40*, 159.
(79) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
(80) Steiner, T. *Acta Cryst. Sec. B* **2000**, *56*, 673.
(81) Anghel, A. T.; Day, G. M.; Price, S. L. *CrystEngComm* **2002**, *4*, 348.
(82) Davey, R. J. *Mol. Cryst. Liq. Cryst.* **1994**, *242*, 79.
(83) Beyer, T.; Lewis, T.; Price, S. L. *CrystEngComm* **2001**, *44*, 1.
(84) Day, G. M.; Price, S. L.; Leslie, M. *Cryst. Growth Des.* **2001**, *1*, 13.
(85) Price, S. L. *Adv. Drug Deliv. Rev.* **2004**, *56*, 301.
(86) Desiraju, G. R. *Nat. Mater.* **2002**, *1*, 77.
(87) Elrod, M. J.; Saykally, R. J. *Chem. Rev.* **1994**, *94*, 1975.
(88) Gavezzotti, A. *Modelling Simul. Mater. Sci. Eng.* **2002**, *10*, R1.
(89) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications: New York, 1996.
(90) Young, D. C. *Computational Chemistry: A practical guide for Applying Techniques to Real-World Problems*, 2001.
(91) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(92) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(93) Mooij, W. T. M.; Leusen, F. J. J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5063.
(94) Stone, A. J. *The theory of intermolecular forces*; Clarendon Press: Oxford, 1996.
(95) Price, S. L.; Stone, A. J.; Alderton, M. *Mol. Phys.* **1984**, *52*, 987.
(96) Price, S. L.; Stone, A. J. *J. Chem. Phys.* **1987**, *86*, 2859.
(97) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
(98) Stone, A. J.; Alderton, M. *Mol.Phys.* **1985**, *56*, 1047.
(99) Price, S. L.; Wibley, K. S. *J.Phys.Chem. A* **1997**, *101*, 2198.
(100) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2004**, *4*, 1327.
(101) Lommerse, J. P., et al. *Acta Cryst. Sec. B* **2000**, *56*, 697.
(102) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872.
(103) Stone, A. J. *Mol. Phys.* **1985**, *56*, 1065.
(104) Lesueur, C. R.; Stone, A. J. *Mol. Phys.* **1994**, *83*, 293.
(105) Axilrod, P. M.; Teller, E. *J. Chem. Phys.* **1943**, *11*, 299.
(106) Barker, J. A.; Fisher, R. A.; Watts, R. O. *Mol. Phys.* **1971**, *21*, 257.
(107) Spackman, M. A. *J. Phys. Chem.* **1989**, *93*, 7594.
(108) Rappe', A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.
(109) Hayes, I. C.; Stone, A. J. *Mol. Phys.* **1984**, *53*, 83.

(110) Hayes, I. C.; Hurst, G. J. B.; Stone, A. J. *Mol. Phys.* **1984**, *53*, 107.
(111) Price, S. L. *CrystEngComm* **2004**, *6*, 344.
(112) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(113) Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2003**, *125*, 16434.
(114) Munowitz, M. G.; Wheeler, G. L.; Colson, S. D. *Mol. Phys.* **1977**, *34*, 1727.
(115) Stone, A. J. *Chem. Phys. Lett.* **1993**, *211*, 101.
(116) Wheatley, R. J.; Meath, W. J. *Mol. Phys.* **1993**, *80*, 25.
(117) Moeller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(118) Hepburn, J.; Scoles, G.; Penco, R. *Chem. Phys. Lett.* **1975**, *36*, 451.
(119) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
(120) Price, S. L.; Andrews, J. S.; Murray, C. W.; Amos, R. D. *J. Am. Chem. Soc.* **1992**, *114*, 8268.
(121) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. *Phys. Rev. Lett.* **2003**, *91*.
(122) Misquitta, A. J.; Szalewicz, K. *Chem. Phys. Lett.* **2002**, *357*, 301.
(123) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.
(124) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(125) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.
(126) Dauberosguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins* **1988**, *4*, 31.
(127) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. *Isr. J. Chem.* **1994**, *34*, 195.
(128) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
(129) Allinger, N. L.; Chen, K.; Lii, J.-H. *J. Comput. Chem.* **1996**, *17*, 642.
(130) Mooij, W. T. M.; Van Eijck, B. P.; Kroon, J. *J. Am. Chem. Soc.* **2000**, *122*, 3500.
(131) Lennard-Jones, J. E. *Proc. Roy. Soc.* **1924**, *106A*, 441.
(132) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J.Am.Chem.Soc.* **1995**, *117*, 5179.
(133) Pertsin, A. J.; Kitaigorodskii, A. I. *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*; Springer Verlag: Berlin, 1987.
(134) Gavezzotti, A.; Filippini, G. *Acta Chim. Hung.* **1993**, *130*, 205.
(135) Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831.
(136) Williams, D. E.; Cox, S. R. *Acta Crystallogr. Sec. B* **1984**, *40*, 404.
(137) Williams, D. E.; Starr, T. L. *Comput. Chem.* **1977**, *1*, 173.
(138) Williams, D. E.; Weller, R. R. *J. Am. Chem. Soc.* **1983**, *105*, 4143.
(139) Williams, D. E. *J. Mol. Struct.* **1999**, *486*, 321.
(140) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1.
(141) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1154.
(142) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.
(143) Brodersen, S.; Wilke, S.; Leusen , F. J. J.; Engel, G. *Phys.Chem. Chem. Phys.* **2003**, *5*, 4923.

(144) Nowell, H.; Price, S. L. *Acta Cryst. Sec. B* **2005**, *61*, 558.
(145) Karamertzanis, P. G.; Price, S. L. *J. Chem. Theory Comput.* **2006**, *2*, 1184.
(146) Dzyabchenko, A. V. *J. Struct. Chem.* **1984**, *25*, 416.
(147) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622.
(148) Verwer, P.; Leusen , F. J. J. *Computer Simulation to predict possible crystal polymorphs*; in K.B. Lipkowitz and D.B. Boyd eds. "*Reviews in Computational Chemistry*"
John Wiley & Sons
1998; Vol. 12.
(149) Williams, D. E. *Acta Cryst. Sec. A* **1980**, *36*, 715.
(150) Gavezzotti, A. *PROMET(5.3) A Program for the generation of possible crystal structures from molecular structure of organic compounds* Milano, Italy, 1999.
(151) Hofmann, D. W. M.; Lengauer, T. *Acta Cryst. Sec. A* **1997**, *53*, 225.
(152) Hofmann, D. W. M.; Lengauer, T. *J. Mol. Model.* **1998**, *4*, 132.
(153) Hofmann, D. W. M.; Lengauer, T. *J. Mol. Struct.* **1999**, *474*, 13.
(154) Tajima, N.; Tsuzuki, S.; Tanabe, K.; Aoki, K.; Hirano, T. *Electron. J. Theor. Chem.* **1997**, *2*, 139.
(155) Arikawa, T.; Tajima, N.; Tsuzuki, S.; Tanabe, K.; Hirano, T. *Theochem-J. Mol. Struct.* **1995**, *339*, 115.
(156) Hirano, T.; Tsuzuki, S.; Tanabe, K.; Tajima, N. *Chem. Lett.* **1995**, 1073.
(157) van Eijck, B. P.; Kroon, J. *J. Comput. Chem.* **1999**, *20*, 799.
(158) van Eijck, B. P.; Kroon, J. *Acta Cryst. Sec. B* **2000**, *56*, 535.
(159) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *J. Am. Chem. Soc.* **2000**, *122*, 3500.
(160) van Eijck, B. P.; Kroon, J. *J. Phys. Chem. B* **1997**, *101*, 1096.
(161) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *J. Comput. Chem.* **2001**, *22*, 805.
(162) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Phys. Chem. A* **1999**, *103*, 9883.
(163) Chaka, A. M.; Zaniewski, R.; Youngs, W.; Tessier, C.; Klopman, G. *Acta Cryst. Sec. B* **1996**, *52*, 165.
(164) Holden, J. R.; Du, Z. Y.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.
(165) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628.
(166) Aakeroy, C. B.; Nieuwenhuyzen, M.; Price, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 8986.
(167) Beyer, T.; Price, S. L. *J. Phys. Chem. B* **2000**, *104*, 2647.
(168) Price, S. L.; Wibley, K. S. *J. Phys. Chem. A* **1997**, *101*, 2198.
(169) Beyer, T.; Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2001**, *123*, 5086.
(170) Karfunkel, H. R.; Gdanitz, R. J. *J. Comp. Chem.* **1992**, *13*, 1171.
(171) Gdanitz, R. J. *Chem. Phys. Lett.* **1992**, *190*, 391.
(172) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(173) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.
(174) Payne, R. S.; Roberts, R. J.; Rowe, R. C.; Docherty, R. *Int. J. Pharm.* **1999**, *177*, 231.
(175) Press, W. H.; Flannery, B. P.; al., e. *Numerical Recepies in C*, 2002.
(176) Della Valle, R. G.; Venuti, E.; Brillante, A.; Girlando, A. *J. Chem. Phys.* **2003**, *118*, 807.
(177) Karamertzanis, P. G.; Pantelides, C. C. *J. Comput. Chem.* **2005**, *26*, 304.
(178) Karamertzanis, P. G.; Pantelides, C. C. *Mol. Simul.* **2004**, *30*, 413.

(179) Schmidt, M. U.; Englert, U. *J. Chem. Soc.-Dalton Trans.* **1996**, 2077.
(180) Pillardy, J.; Czaplewski, C.; Wedemeyer, W. J.; Scheraga, H. A. *Helv. Chim. Acta* **2000**, *83*, 2214.
(181) Pillardy, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 12351.
(182) Williams, D. E. *Acta Cryst. Sec. A* **1996**, *52*, 326.
(183) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C.; Scheraga, H. A. *J. Am. Chem. Soc.* **2000**, *122*, 907.
(184) Wawak, R. J.; Pillardy, J.; Liwo, A.; Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem. A* **1998**, *102*, 2904.
(185) Bazterra, V. E.; Ferraro, M. B.; Facelli, J. C. *J. Chem. Phys.* **2002**, *116*, 5984.
(186) Payne, R. S.; Rowe, R. C.; Roberts, R. J.; Charlton, M. H.; Docherty, R. *J. Comput. Chem.* **1999**, *20*, 262.
(187) Leusen , F. J. J. *J. Cryst. Growth* **1996**, *166*, 900.
(188) Payne, R. S.; Roberts, R. J.; Rowe, R. C.; Docherty, R. *J. Comput. Chem.* **1998**, *19*, 1.
(189) Ouvrard, C.; Price, S. L. *Cryst. Growth Des.* **2004**, *4*, 1119.
(190) Yu, L.; Reutzel-Edens, S. M.; Mitchell, C. A. *Org. Process Res. Dev.* **2000**, *4*, 396.
(191) Lewis, T. C.; Tocher, D. A.; Price, S. L. *Cryst. Growth Des.* **2004**, *4*, 979.
(192) Price, S. L.; Patel, B.; Pridhanani-Jethani, P.; Torrisi, A. *ACA Trans.* **2004**, *39*, 2.
(193) van Ejick, B. P. *Acta Cryst. Sec. B* **2005**, *61*, 528.
(194) Dzyabchenko, A. *Acta Cryst. Sec. B* **1994**, *50*, 414.
(195) Chisholm, J. A.; Motherwell, S. *J. Appl. Crystallogr.* **2005**, *38*, 228.
(196) Barr, G.; Dong, W.; Gilmore, C. J. *J. Appl. Crystallogr.* **2004**, *37*, 658.
(197) Barr, G.; Dong, W.; Gilmore, C. J. *J. Appl. Crystallogr.* **2004**, *37*, 243.
(198) Gavezzotti, A. *CrystEngComm* **2003**, *5*, 429.
(199) Gavezzotti, A. *CrystEngComm* **2003**, *5*, 439.
(200) Gavezzotti, A.; Filippini, G.; Kroon, J.; vanEijck, B. P.; Klewinghaus, P. *Chem.-Eur. J.* **1997**, *3*, 893.
(201) Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L. *J. Phys. Chem. B* **2006**, *110*, 3323.
(202) Coombes, D. S.; Catlow, C. R. A.; Gale, J. D.; Hardy, M. J.; Saunders, M. R. *J. Pharm. Sci.* **2002**, *91*, 1652.
(203) Hulme, A. T.; Price, S. L.; Tocher, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 1116.
(204) Lancaster, R. W.; Karamertzanis, P. G.; Hulme, A. T.; Tocher, D. A.; Covey, D. F.; Price, S. L. *Chem. Comm.* **2006**, 4921.
(205) Timofeeva, T. V.; Kinnibrugh, T.; Borbulevych, O. Y.; Averkiev, B. B.; Nesterov, V. N.; Sloan, A.; Antipin, M. Y. *Cryst. Growth Des.* **2004**, *4*, 1265.
(206) Sarma, J.; Desiraju, G. R. *Cryst. Growth Des.* **2002**, *2*, 93.

Chapter 2

PHASE TRANSITIONS in MOLECULAR CRYSTALS

2.1 Phase Transitions from a molecular point of view

2.1.1 Classification of phase transitions

A crystal can be viewed as a supra-molecule with macroscopic dimensions, containing millions of interacting molecules, held together in a periodic arrangement[1]. It has been observed that many compounds can exist in several solid forms, depending upon factors such as temperature, i.e. one form of a compound can be stable at room temperature, but metastable at another temperature. In the organic solid state the free energy difference between different forms are often small - a few $kJmol^{-1}$ -. A metastable form can persist for years in its form, or it can undergo spontaneous transformation to the more stable one. From a supra-molecular point of view the transformation of one conformation to another can be thought of as a reaction between the supra-molecular species, similar to a chemical reaction.

Let us now examine the types of phase transitions[1]:

- *1st order phase transitions* are characterised by a discontinuous change in the first derivatives of the free energy of the system (volume or entropy), at a transition temperature T_t (figure 2.1a). These transitions often occur through heterogeneous nucleation and growth processes. The mechanism of formation of new phases is generally denoted *reconstructive*, and this leads to major differences in the structure of the new phase with respect to its starting structure. These processes usually occur at the interface between the newly forming and the old phase. There is no orientational relationship between the two phases. The rate of transition is a function of temperature only, but hysteresis is always present. An example of 1st order phase transition is represented by the conformational polymorphic transition in the dimethyl ester, 2,5 dihydroxy 3,6 dichloroterephthalic acid[1].

- *2nd order phase transitions* are characterised by a change in the gradient of a first derivative property (volume or entropy) with temperature, but without discontinuity at the transition temperature (figure 2.2b). These transitions are believed to occur through a "softening" of normal modes of vibrations in the lattice, which fall to zero at the transition temperature, and cause a correlated displacement of the atoms in the crystal; no hysteresis is observed in this type of transition. 2nd order phase transitions are rapid, with the rate not depending on temperature. We will describe these transitions later in this section. 2nd order phase transitions include inorganic and organic solid state *displacive* and *martensitic* transitions[2-7].

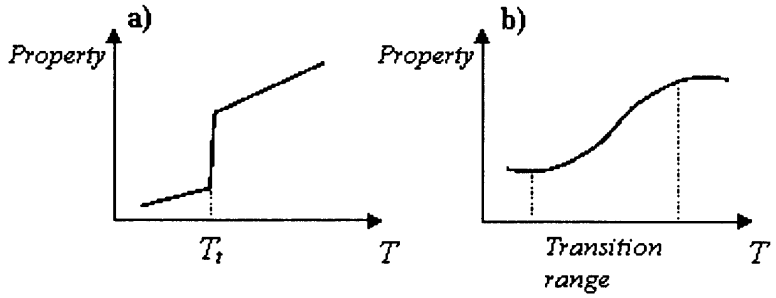


Fig.2.1: First derivative generic property as a function of temperature in the two types of phase transition: **a)** 1st order phase transition. **b)** 2nd order phase transition.

The noted classification is quite general and often the distinction between 1st and 2nd order transitions is blurred. Indeed, Mnyukh criticises[8] this classification and considers all phase transitions as *non*-equilibrium phenomena, in which the generation of a new phase passes through a necessary superheating or supercooling condition. At the transition temperature, the two phases are in equilibrium and resulting in no change at this stage of the process. From this perspective, all phase transitions occur through a mechanism of nucleation and growth, in which the former is the key step and depends on the presence of defects, i.e. micro-cavities, and surface irregularities between the two crystal domains. Mnyukh also claims that, as a general rule, there is a total absence of orientational relationship between the "parental" and the "daughter" phase[9]. Nevertheless, the factors which govern nucleation and growth are still not understood and the orientational relationship is sometimes observed and sometimes not.

We might describe a phase transition between two polymorphs as a simple deformation of one structure into the other, which involves translations and rotations of the individual molecules into another state. Such a description, however, can be applied only in some

cases: when the unit cells of the two phases are quite similar. In general, it seems more likely that the molecules of a "parental" phase do not have high mobility within the existing phase and they are transferred individually to the new stable phase across phase boundaries[1].

2.1.2 Phase Transitions & Solid State Reactions

Another important phenomenon, which is also observed in molecular crystals, are solid-state reactions[1]. A distinction has to be done between solid-state reactions and phase transitions, but the distinction becomes more difficult for structural transitions involving changes of hydrogen bonds patterns as well as molecular conformations.

In general any mono-molecular reaction, which takes place in solution can occur also in the solid phase, but it will have different rates of reaction in the two phases, due to the different environment surrounding a molecular species in the solid phase compare to that in solution. *Topochemical* reactions is the term used for solid-state reactions, which are more favourable in the solid state than in solution, by taking advantage of the packing arrangement of the atoms in the crystal[10], exploiting a minimal movement of atoms or molecules in the crystal[10]. In the early stages of a solid-state reaction the product accumulates as a solid solution in the "parental" phase before undergoing nucleation and precipitate into a new separate phase. An example of such transformations is given by the methyl transfer reaction in *p*-N,N-dimethylaminobenzenesulphonic acid methyl ester[11], in which upon warming, an intermolecular migration of a methyl group occurs, which yields the zwitteranion form of the molecule. In the zitteranion form, methyl migration can occur with minimal atomic motion, within the short distance between ordered molecular arrangements, allowing this reaction to be possible. Another example of a *topochemical* reaction includes proton transfer reactions, occurring in many heterocyclic crystals[12], in particular imidazole[13] and the imidazole-based compounds, which we will discuss later in this thesis (section 4.5).

2.1.3 Order - Disorder phase transitions: phase transition in terephthalic acid

Order-Disorder phase transitions are widely observed in organic molecular crystals. It is important first to define the nature of the disorder. Indeed the kind of disorder occurring in these transitions is not simply a static disorder in atomic position typical of amorphous glasses, but often involves a disorder in the stacking of chains or layers of interacting molecules to generate a 3D structure. This is the case for *p*-N,N-

dimethylaminobenzenesulphonic acid methyl ester[11]. In other cases, however, disorder relates to the disorder of the hydrogen positions, which is very common in crystals with intermolecular dimers like terephthalic acid[14]. This compound has two crystal structures, form I and form II, both of which characterised by molecular arrangements in double hydrogen-bonded chains, which run parallel along planes.

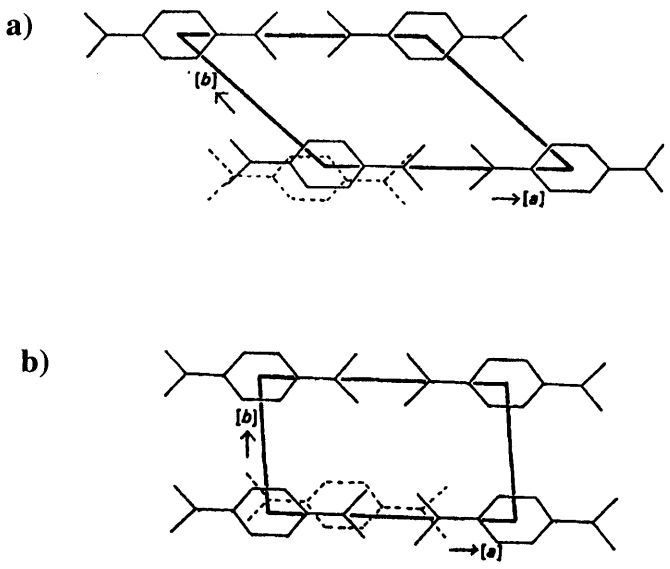


Fig. 2.2: The relationship between the packing type of two polymorphs of terephthalic acid along ***c*** axis: **a)** form 1 and **b)** form II. Dashed lines represent the molecule lying immediately above that, at the origin[14].

The two structures, which are both in triclinic form, differ in the stacking of the chains along the planes and in the stacking of the planes in forming a 3D structure. In form I neighbouring chains pack with benzene rings of one chain adjacent to the carboxylic groups of the next chain, whereas in form II the benzene rings of adjacent chains are in line. The opposite situation is observed in the stacking of the layers. Neutron scattering studies[15-17] have revealed the presence of disorder in the acid hydrogens of the triclinic form I, stable at T > 373K. The proton disorder comes from the dimeric hydrogen bond interaction in the structure observed in many carboxylic acids[18], which in the terephthalic acid crystal structure has two tautomeric forms (figure 2.3):

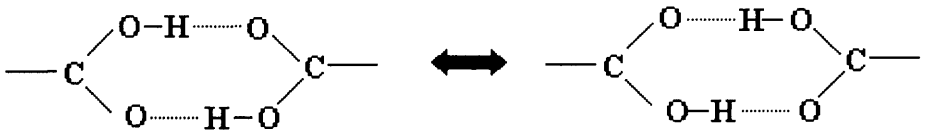


Fig. 2.3: Tautomeric dimers in carboxylic acids: i.e. terephtalic acid.

The two tautomers in figure 2.3 are related to each other by a double proton transfer mechanism. Further neutron scattering studies revealed a dynamical behaviour solely governed by a simple classical two-site exchange process[15], with a small activation energy of ~ 2.6 kJ mol^{-1}. In the stable structure at low temperature (form II), the proton is localised on a particular tautomer, revealing an asymmetric nature in the double well potential energy of the system[17,19], with the two tautomers not being energetically equivalent[17]. The proton exchange mechanism is very fast[20] and generates disorder in the high temperature structure of form I. More recent X-ray investigations have established the stability of form II at room temperature and pressure[19]and interpreted the phase transition ordered form II to disordered form I to be driven by the contribution of the proton disorder to the Gibbs free energy change.

The kind of order-disorder transitions, illustrated above (figure 2.2), may sometimes involve breaking of hydrogen bonds, which require overcoming high-energy barriers and sometimes can be occur with *reconstructive* mechanism.

A further kind of order-disorder transition refers to the disorder in the molecular orientations of a crystal structure at high temperature, with molecules subjected to freedom in rotational movement around their centre of masses. This kind of disorder is treated thoroughly in section 2.5.

2.2 "Soft Mode" Phase Transitions

2.2.1 Anharmonic Effects

A large group of phase transitions, previously classified in the group of "2nd order phase transitions", are strictly related to the anharmonic interactions of phonons in crystals[21]. In the harmonic approximation of phonon interactions, the Hamiltonian of the system can be expressed as:

$$H = \frac{1}{2}\sum_{k,\nu} \dot{Q}(\boldsymbol{k},\nu)\dot{Q}(-\boldsymbol{k},\nu) + \frac{1}{2}\sum_{k,\nu} \omega_0^{\,2} Q(\boldsymbol{k},\nu)Q(-\boldsymbol{k},\nu) \quad , \qquad [2.1]$$

where $Q(\boldsymbol{k},\nu)$ is a generic normal mode coordinate, $\boldsymbol{k}$ is the wave vector of the reciprocal lattice and ω_o is the harmonic phonon frequency of the mode ($\boldsymbol{k}$, ν). The form of the potential is parabolic (figure 2.4a). Anharmonic effects include other terms in the Hamiltonian of the system [2.1], which are high order phonon-phonon interactions. The quasi-harmonic approximation considers interactions up to the quartic term (4th body interaction). In the latter approximation the Hamiltonian has still the expression of the

harmonic approximation [2.1], but with a "renormalized" phonon frequency $\varpi(\boldsymbol{k},\nu)$ which is related to the harmonic frequency by the relationship:

$$\varpi^2 = \omega_0^{\,2} + \alpha T \qquad [2.2]$$

The quasi-harmonic approximation considers the coupling constant to be independent on the temperature and positive ($\alpha > 0$). The effect of the quartic term in the quasi-harmonic approximation is to shift the potential with respect to the harmonic approximation, whose parabolic form is maintained, but is this time temperature dependent (figure 2.4b). As the temperature increases, the potential becomes narrower than the harmonic one and the atoms are constrained to a tighter amplitude, which gives rise to a higher frequency. The inclusion of anharmonic terms (> 4) would result in the appearance of large deviations in the potential from the parabolic form (figure 2.4c).

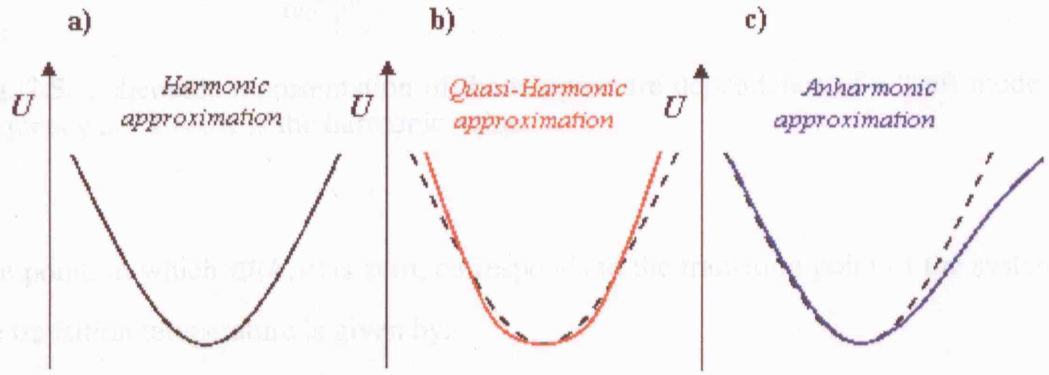


Fig. 2.4: Schematic representation of the potential well for a single atom within the different approximations. In b) and c) the dashed curve represents the potential in the harmonic approximation.

The picture of a phonon frequency, which varies linearly with temperature, can explain the nature of so called *displacive* or "soft mode" phase transitions. A crystal in which the interacting phonons are described within the harmonic approximation, becomes unstable against the displacement of the atoms, which correspond to a negative harmonic frequency mode ω_0^2, or to an imaginary phonon frequency ω_0. In the quasi-harmonic approximation this would imply a transition to another more stable phase with a coupled change in the temperature. In fact, starting from a high temperature symmetric phase, if one imaginary normal mode frequency at any value of $\boldsymbol{k}$ is present, then the structure will not be stable at 0K ($\varpi^2 = \omega_0^2$) and it will transform into another more stable structure with lower symmetry via a distortion, or, more precisely, an atomic displacement, whose

associated imaginary frequency is ω_o. The evolved low-temperature structure will be equivalent to the high symmetric one, with one frozen-in normal mode, corresponding to the one with an imaginary frequency. This mode is called a "soft mode", because it has a low frequency and the crystal is essentially "soft" against the corresponding atomic displacement associated with the mode. Upon heating anharmonic contributions increase until the renormalized frequency ($\varpi^2 = \omega_0^2$) becomes zero, and then real (figure 2.5). At this point of the transition the high symmetry structure is stable and the low-temperature, low-symmetry structure relaxes to the high-symmetry one.

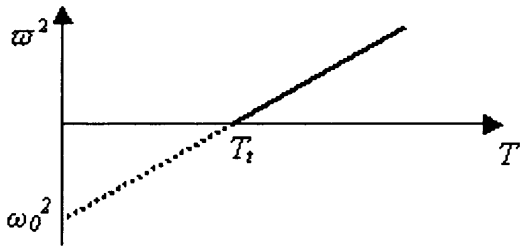


Fig. 2.5: Schematic representation of the temperature dependence of a "soft mode". The frequency at T = 0K is the harmonic value.

The point, at which $\varpi(k,\nu)$ is zero, corresponds to the transition point of the system, and the transition temperature is given by:

$$T_t = -\frac{\omega_0^2}{\alpha} \quad . \qquad [2.3]$$

The anharmonic effects, which yield displacive phase transitions, are not particularly strong. Examples of such kinds of transitions are observed in a number of molecular crystals, such as 1,2,4,5 tetrachlorobenzene[21,22], and s-triazine[23].

2.2.2 Ferroelectric or Zero-centre phase transitions

Here we discuss another type of "soft mode" phase transition, which yields a macroscopic dielectric polarisation of the whole crystal. Ferroelectric phase transitions, illustrated in figure 2.6, involve the loss of a centre of symmetry in the unit cell, which produces a net dipole moment in the cell.

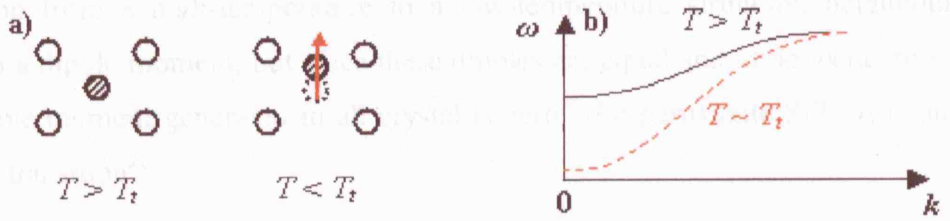


Fig.2.6: a) Schematic atomic displacement in the ferroelectric "soft mode" phase transition. **b)** schematic modification of the optical transverse mode "softening" with temperature in the ferroelectric phase transition.

Ferroelectric transitions have a tendency to shift the dielectric constant to an infinite value for $T \sim T_t$ (transition temperature). This effect results in a softening of a transverse optical normal mode with $\boldsymbol{k} \sim 0$, as shown in figure 2.6b. Such phase transitions are observed in a large number of perovskites[24] and in the oxide $PbTiO_3$ [25].

2.2.3 Antiferroelectric or zone boundary phase transition

Antiferroelectric transitions are associated with a "soft mode", with wave vector, $\boldsymbol{k}$, at the Brillouin zone boundaries. In this type of transition the "soft mode" can be either optical or acoustic and the distinction between the two types (i.e. optical and acoustic) is not so clear at the phase boundaries[21]. The softening is illustrated in figure 2.7a and 2.7b.

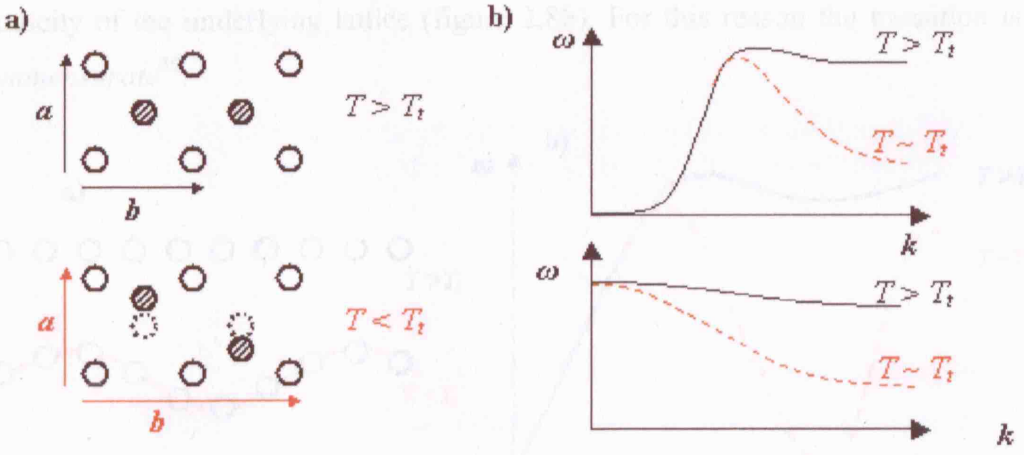


Fig. 2.7: a) Schematic atomic displacement in the antiferroelectric "soft mode" phase transition. **b)** schematic modification of the acoustic (upper graph) and optical (lower graph) transverse mode "softening" with temperature in the antiferroelectric phase transition.

One characteristic of the antiferroelectric phase transformation is the doubling in one or more directions of the unit cell of the low-temperature phase (figure 2.7a). In the

transition from a high-temperature to a low-temperature structure, neighbouring cells develop a dipole moment, but since these dipoles are equal and in opposite directions, the net dipole moment generated in all crystal is zero. The perovskite $SrTiO_3$ is an example of such transition[26].

2.2.4 Ferroelastic phase transition

Ferroelectric phase transitions are analogous to *ferroelectric* phase transitions, in which a reversible spontaneous shear strain is generated at temperatures close to the transition temperature T_t. Hysteresis behaviour is observed. The phase transition is accompanied by a combination of elastic constants, which falls to zero at T_t and the soft mode is usually an acoustic normal mode. Such kind of transition is observed in sym-triazine[27] and in HCN. For the latter, a study of the deuterated form (DCN) was performed with neutron scattering[28,29].

2.2.5 Incommensurate phase transition

Incommensurate phase transitions are determined by the occurrence of a "soft mode" at a certain value of $\boldsymbol{k}$, which is situated between the zone centre and the zone boundary (figure 2.8a). The accompanying "soft mode" yields a distortion in the atomic displacement, which imposes a periodicity on the crystal structure unrelated to the periodicity of the underlying lattice (figure 2.8b). For this reason the transition is called *incommensurate*[30].

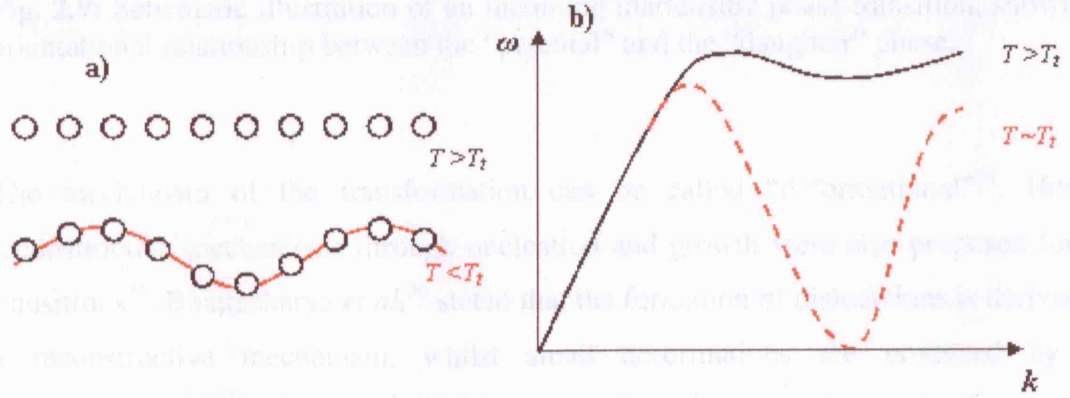


Fig. 2.8: a) Schematic atomic displacement in the incommensurate "soft mode" phase transition. **b)** schematic modification of the acoustic transverse mode "softening" with temperature in the incommensurate phase transition.

There are various mechanisms for *incommensurate* phase transitions, the most common of which is given by a subsequent "lock-in" transition. Lock-in transitions occur at lower degrees under the transition temperature T_t and relate the periodicity of the modulation changes to the periodicity of the crystal lattice. An example of *incommensurate* phase transition is given by that undergone by the oxide K_2SeO_4 [30].

2.2.6 Martensitic phase transitions

Martensitic phase transitions are "soft mode" transitions, in which an orientational relationship is observed between the "parental" and the emerging "daughter" crystals, also called topotaxy[31]. These transitions are very fast, with velocity in the order of the speed of sound (344 m s^{-1}) and it is often characterised by a cooperative behaviour of molecules in the bulk crystal[32]. The term "martensitic" implies that under an appropriate deformation (shear), molecules fall into their sites in the new phase through small displacements, which do not require dislocations[33] and transformation twinning. (figure 2.9).

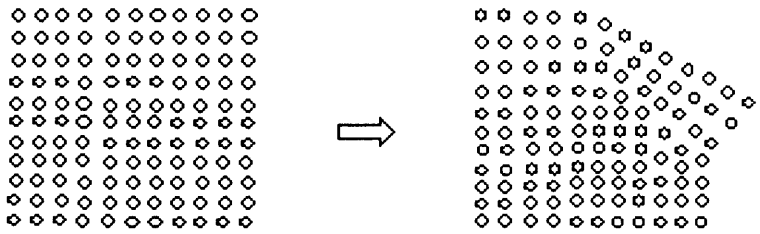


Fig. 2.9: Schematic illustration of an incoming martensitic phase transition, showing the orientational relationship between the "parental" and the "daughter" phase.

The mechanism of the transformation can be called "deformational"[34]. However, reconstructive mechanisms through nucleation and growth were also proposed for these transitions[35]. Bhattacharya *et al.*[36] stated that the formation of dislocations is derived from a reconstructive mechanism, whilst small deformations are produced by weak transformations (figure 2.10). In the presence of significant dislocations, the martensitic transition will not be reversible, while on the contrary, it would be fully reversible if occurring via small deformations.

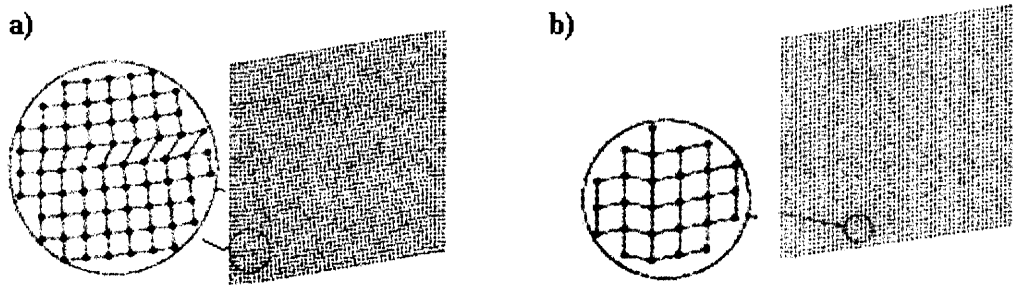


Fig. 2.10: a) Reconstructive phase transformation: the region highlighted shows the occurrence of a dislocation. **b)** the small deformations in the lattice, produced by weak transformations. Figure from Bhattacharya *et al.* [36]

There are few examples of organic molecular crystals exhibiting this kind of transition. An example is *p*-di-iodochlorobenzene[37], for which the martenisitic phase transition is first order. The same phase transformation has been observed in petroselinic acid[38], under rapid cooling of the high melting monoclinic form. Another example is given by D,L Norleucine[35]. A theoretical study of the latter has been performed with molecular dynamics simulation, with a careful derivation of the intermolecular potential for the two phases[32]. The simulation reproduces the transformation from form α, in which packing of bi-layer units is of the form A-A-A-A, to form β, with an alternate packing of bi-layer units A-B-A-B, by half cell length along ***b*** axis. The two forms have similar free energies, with a difference of ~ 1.7kJmol^{-1}. The behaviour of the phase transition depends on a delicate balance between enthalpic and entropic differences for the two crystalline phases. A key issue is the nature of the molecular processes at the interfaces, which were not observed in the simulation. The transformation seems to occur very sharply, beginning by the de-stabilisation of the van der Waals surface between the bi-layers, with some molecule shifting along the b axis. The packing of these molecules oscillates at interfaces between the two phases. These oscillations might be an artefact of the small system size, which gives very large fluctuations in the simulation cells. However, the system size is always a critical factor in molecular dynamic simulations. Since we are limited by the computational cost of molecular dynamic simulation, we are not yet usually able to simulate *reconstructive* phase transitions with a mechanism of nucleation and growth.

2.3 Plastic Crystals

2.3.1 Plastic Crystals: properties and applications

Liquid and plastic crystals, also referred to as mesophases, represent a number of different states of matter, in which the degree of molecular order lies between the almost perfect crystalline long-range positional and orientational order and the statistical long-range disorder(typical of amorphous liquids and gases).

The first liquid crystal recognised was in 1888 by Reinitzer[39], who was investigating cholesteryl benzoate. He noted that the substance melted to a cloudy liquid at 145.5°C and became a clear liquid at 178.5°C. Interestingly, he also observed that, upon cooling, the liquid briefly turned blue at the transition temperature, and that a blue violet colour appeared just before crystallization. He first pointed out the presence of two melting points, and was the first to use the name "liquid crystals" for this intermediate phase, a term that is now generically accepted. Many liquid crystalline substances are organic molecules consisting of a long non-polar chain and a polar "head". Commonly they also contain aromatic rings. Liquid crystals have also been known to be disc-like (discotic) in shape.

A very extensive series of mesophases, first recognised in 1935 by Timmermans[40], are referred to as plastic crystals[41]. They are amongst the most disordered mesomorphic phases and the most "liquid" in character. They separate at relatively high temperature from low temperature perfect crystalline phases, often described by simple cubic symmetry and have high plasticity, in terms of plastic deformation. These compounds are neither true liquids (because of conserving long-range order), nor true crystalline solids (as they have a complete loss of long-range orientational, and have partial positional order[42]). They constitute a further mesomorphic state of matter. Plastic crystals are typically constituted by compact "globular" molecules[43], which are not strongly aniso-dimensional and which acquire "close-to-spherical" symmetry when undergoing rotatory displacements. Usually, these mesophases have a very low entropy of liquefaction ($\Delta S_m <$ 0.076 $kJmol^{-1}K^{-1}$) and their formation over a certain range of temperatures is due to the capacity of the constituent molecules to arrange themselves in a very high symmetric (often cubic, sometimes hexagonal) array, whilst undergoing simultaneous thermal rotations, in no long-range orientational order amongst the molecules. In some phases, the free rotations and the tumbling motion of the molecules occurs more freely even than in the liquid phase[44]. It has been shown that at the upper limit of the range of temperature,

in which the plastic phase is stable, liquefaction occurs, with a breakdown of the high symmetric arrangement of molecules, but with an associated small change in entropy ΔS (for cyclopentane[42] $\Delta S_m \sim 0.012$ kJmol^{-1}K^{-1}). On the other hand, at the lower temperature limit, a transition from plastic crystal I – ordered crystal II occurs, typically to an ordered anisotropic crystal structure. This transition is accompanied by a large decrease in entropy (see figure 2.11).

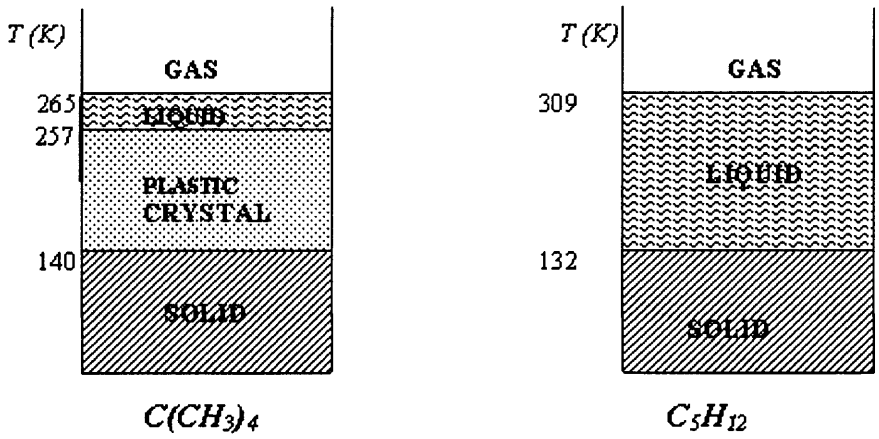


Fig. 2.11: Schemes of the transition phases for neopentane, $C(CH_3)_4$, with "globular"shape and for pentane, C_5H_{12}, with *non*-"globular" shape.

Occasionally there are transitions to a second and a third plastic phase, as in the case of cyclopentane[45], which are associated with a small change in entropy, and are followed by a final low temperature transition to an ordered anisotropic solid phase, with a large entropy variation; successive plastic phases likely differ in details of the permissible thermal displacement of the molecules, a phenomenon that has yet to be fully investigated.

For a molecule to exhibit a plastic phase it needs to possess high symmetry and globularity. Not all globular and highly symmetric molecules crystallise in a plastic phase. For example, methane, silane and carbon tetrachloride all form a plastic phase, whilst silicon tetrachloride not. Analogously, adantamane forms a plastic phase, whereas hexamethyltetramine, which can be considered as an adantamane in which the C-H groups are replaced by the nitrogen (N), forms an ordered *non*-rotatory body centred cubic crystal. This could be due to the presence of polar groups in the molecules (as *N* atoms in hexamethyltetramine) or due to an elongation of bonds (i.e. the Si-Cl bond in silicon tetrachloride), which may hinder atomic re-orientations. An extensive review about this kind of compounds is given in Sherwood's book[46].

Let us consider some of the physical properties exhibited by plastic crystals.

- *Thermal properties:* as previously stated, plastic crystals typically have a small entropy and high temperature of melting. In some cases they can directly sublime, without passing from the melt as in perfluorocyclohexano and hexafluorosulfide[42]. Solid crystal – plastic crystal transitions are usually accompanied by a high entropy increment and very similar values of the entropy in the plastic crystal and liquid phase underline the high molecular disorder present in the former case. Such a disorder is orientational and the number of permissible molecular orientations randomly adopted by the molecules in the plastic phase, can be estimated from measurements of heat capacities and entropies during the phase change[43]. Finally, the thermal expansion of the plastic phase is larger than that of the corresponding ordered solid phase[47].
- *X-ray scattering*: the relative high degree of disorder in plastic phases is reflected in X-ray powder patterns by a small number of reflections, with considerable background scattering[48]. Plastic crystals have an interesting general characteristic: the symmetry elements of the unit cell are incompatible with those of the molecules, if the latter are regarded as static[49]. Such discrepancies between the lattice and molecular symmetry of a plastic crystal provide evidence of the presence of high disorder in the molecular orientation. As previously highlighted, the plastic crystals usually have cubic face centred (f.c.c.), or body centred (b.c.c.) unite cells. Occasionally, however, the presence of hexagonal unit cells has been observed[50] and, if the average of the molecular orientations is close to having spherical symmetry, close packing (hcp or ccp) cells may be expected for the spherically (or isotropically) rotating molecules in the plastic phases of molecular crystals.
- *Dielectric Properties*: several studies of the dielectric properties of plastic crystals have been performed for "globular" molecules possessing a dipole moment, such as andantamanone, pentachloronitrobenzene, cyclo-hexanol, ethanol, meta-carborane and cyano-adantamane[51]. In the crystal structures of these compounds, which do not crystallise in a plastic phase, a sharp drop in the dielectric constant, ε, is observed, during the liquid - crystalline transition (see figure 2.12).

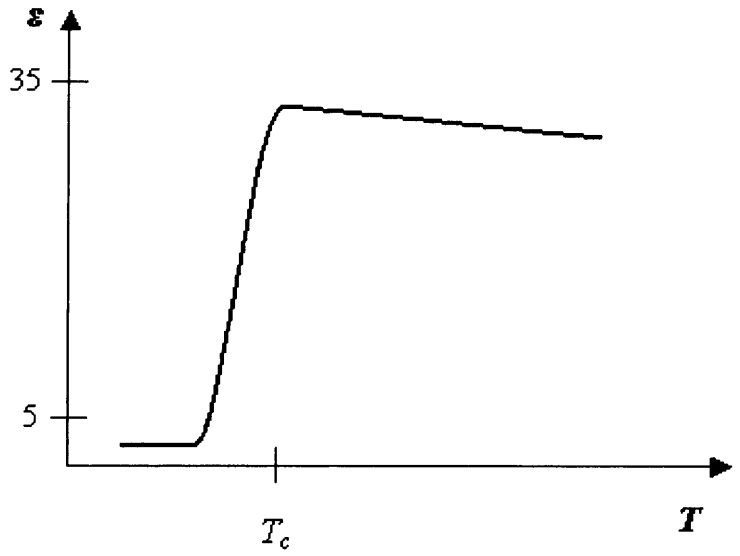


Fig. 2.12: Variation of the dielectric constant, ε, as a function of temperature for a solid ordered crystalline phase.

The variation in the dielectric constant can be explained by the presence of orientational polarizability of the compound in an electric field, which drops to zero when the molecular dipoles become fixed in an ordered low temperature crystalline lattice[52]. In the case of plastic crystals, on account of rotational re-orientation of the molecules, the dielectric constant does not fall to zero, but retains a value of the same order as that for the liquid[42]. A further sharp drop occurs in the plastic crystal – solid ordered crystal transition.

Measurements of the dielectric constants give useful insights into the rotational characteristics of the plastic phase for a number of compounds, such as tetrasubstituted methane[52]. Dielectric relaxation has been previously studied for a large number of plastic crystals to establish structural properties as fragility at different temperatures[51], such as plastic crystals Freon 112 and Freon 112a[53]. However, for some compounds, forming plastic phases leads a decrease in the internal viscosity or resistance with respect to the amorphous liquid state[44]. Generally, decreasing the temperature of plastic phases raises their resistance to molecular rotational motion.

- *NMR spectroscopy*: NMR studies of plastic phases have been performed to investigate self-diffusion and molecular rotation[43,45,54,55,56]. However this method is limited since it cannot distinguish between literally free rotation of the molecules and their capability to tumble rapidly from one orientation to another[42].

Plastic crystals containing more than one type of molecular unit are also known[43]. Compounds forming plastic crystals are frequently miscible in the solid plastic phase[57-60], as in the case of methane – krypton or cyclopentane – 2,2 dimethylbutane. However, such behaviour is not a rule and indeed, cyclohexane and carbontetrachloride, both of which have plastic phases, are not miscible[57]. Compounds giving two- unit plastic phases are quite common among inorganic compounds such as ammonium nitrate $(NH_4)NO_3$[61]. Often these solid solutions are accompanied by high diffusion of a variety of species in the plastic phase, which can give unique electrical conductivity to the material.

There are several different applications for plastic crystals, which take advantage of their unique properties. Two important cases are worthy of note:

- *Energy Storage*: Compounds such as pentaglycerin and neopentylglycol are promising in the application of energy storage[62,63] in the plastic phase transformation either as pure compounds or as solid solutions of the two[64-67].
- *Solid State Ionic Conducion*: Given the highly disordered nature of plastic crystal, they can give an enhanced diffusivity to ionic species introduced or already present in their matrix[68], as in the case of organic salts based on quaternary ammonium-like species[69]. When doped with salt of acronym, LiTFSI[69], with formula $Li(CF_3\ SO_2)_2N$, the conductivity reaches a high value, about $2 \cdot 10^4$ Scm^{-1} at room temperature. Studies of non-ionic highly polar plastic phase of succinonitrile opportunely doped with ionic species have demonstrated this particular plastic crystal to be very promising as a fast ionic conductor[70,71]. Promising plastic phase ionic conductors are also based on pyrazolium imides[72,73].

2.3.2 Plastic Crystals: types molecular motions

There are two main molecular motions exhibited by plastic phases:

1) *Molecular rotations*: the energy barrier to molecular rotation in plastic crystals is so small that molecules can be thought of as rotating in a literal sense around their centre of mass. The barrier between minima in the potential energy is small enough to allow the molecule to switch from one orientation to another quite rapidly. Hence, in a plastic phase, molecules are distributed over a wide range of different possible orientations[43].

   From the van der Waals radii of spherical atoms of a molecule, it is possible to estimate the volume of a sphere, which would circumscribe a freely rotating molecule. In a plastic crystal, this volume is always bigger than the distances

between the two centres of mass of two neighbouring molecules[49,74]. This fact prevents molecules from undergoing fully free isotropic rotational dynamics, as they do not have enough free volume available for such a motion. Hence, changes in the orientations of molecules require correlated movements among neighbouring molecules.

The volume change on melting is generally small, and even in the liquid phase, there is insufficient space for free molecular rotations. Information about rotational movements can be provided by the analysis of the heat capacity at constant pressure, C_p, of the plastic phase[75]. From this quantity the heat capacity at constant volume, C_v, can be derived from:

$$C_p = C_v + R \quad . \qquad [2.4]$$

The latter is given by the sum of different contributions, from the lattice vibrations (C_{lat}), from the intramolecular vibrations (C_{int}) and the librational and rotational movements of the molecule (C_{rot}). If the former two contributions can be assessed, C_{rot} for the plastic phase can be obtained as:

$$C_{rot} = C_v - (C_{lat} + C_{int}) \quad . \qquad [2.5]$$

The value of C_{rot} for free rotations is ~ $3/2R$ (with R being the universal constant of ideal gasses) and in general, in the case of plastic crystal it is found to be less than this value[75], indicating that the molecular motion in the plastic phases moves away from being a pure free rotation. There are cases such as methane[76], however, in which the plastic phase, near the melting point, has C_v ~ $3/2R$ and it represents one of the few cases of plastic crystals, in which the energy barrier to rotation at high temperature is small enough to permit virtually free rotations.

Resistance to the re-orientational mechanism in the plastic phases of polar molecules can be derived from measurement of the dielectric constants and dielectric loss[77]. For molecules with internal flexibility in the molecular ring (such as cyclopentane), a coupling correlation is observed between a fast internal puckering motion of the ring, and a slower tumbling re-orientational motion of the molecule, with the internal flexibility affecting the dynamical properties of the system[78]. Nevertheless, this is not a general rule and other flexible cycloalkanes (for instance cyclohexane), although having large internal motions in the molecules in the plastic phase, show rotational dynamics of the structure that are essentially independent of the molecular internal modes[79].

2) *Self-Diffusion*: NMR studies have detected residual interactions arising from non-isotropic motion of protons for different molecules in the plastic crystalline phase[43]. Frequently, when the melting point is approached, the relative motion of hydrogen is found to become isotropic and a self–diffusion process is assumed to take place[80]. NMR studies have also shown that a solid crystal – plastic crystal transitions causes an increase in the oriental freedom of the system, which is not immediately translated into the ability of the molecules to diffuse[57]. In this case, the phase transition is 2nd order and the heat capacity above the transition temperature is almost the continuation of that below. This kind of behaviour is observed in the plastic phase II – phase I transition of hexamethyl-disilane[81], for which an activation energy of the self – diffusion process in hexamethyl-disilane has been estimated to be ~ 48 kJmol^{-1}. Defects like vacancies can influence the process of self-diffusion[49] and shift the temperature of solid - plastic phase transition [82].

2.4 Order-Disorder Phase Transitions in Plastic Crystals

The orientational order-disorder phase transition in molecular crystals has been studied from many points of view. Some of these studies try to give an interpretation of the entropy changes, ΔS_t, associated with the transition. The change of the entropy derives mainly from two important factors:

- the increased number of distinguishable orientations available to the molecules in the disordered plastic phase; which is generally dominant.
- the increased freedom of movement for the molecules in the system, which is associated with a certain degree of self-diffusion.

2.4.1 The Guthrie- McCullough's approach

An attempt to describe the orientational order-disorder phase transition is given by the approach adopted by Guthrie and McCullough[83], which is mainly based on considerations of symmetry and steric factors (in conjunction with thermal data) to estimate the number of possible orientations available to molecules in transforming from an ordered to a disordered plastic phase. This methodology has been applied to systems such CCl_4, which has changes in entropy of ΔS_t ~ 0.075 Jmol^{-1}K^{-1}. The molecular orientations, which will be energetically favoured, are those for which there is some correspondence between the symmetry elements of the molecule and the symmetry elements of the lattice. For

example, for a tetrahedral molecule in a cubic lattice, one possible set of orientations involves the coincidence of the three-fold axis of the molecular tetrahedron with the three-fold axis of the unit cell of the plastic crystal. The increase in the entropy is evaluated from the number of the available orientations in the plastic phase, *Z*, following the formula from the statistical thermodynamics[84]:

$$S = R \cdot lnZ \quad . \qquad [2.6]$$

For the case of carbon tetrachloride, there are ten overall symmetry favoured available orientations in the plastic phase, with a net entropy increment of ~ 0.07 $Jmol^{-1}K^{-1}$, (calculated using equation [2.6]), whilst there are other sets of available orientations, which are *priori* possibilities, and which are not easy to clearly identify; when there is also the presence of conformational disorder in the molecules, this can increases further the entropy increment of the phase transition.

2.4.2 The Pople – Karasz model

The model developed by Pople and Karasz[85,86] to describe the fusion of molecular crystals takes into account the orientational and positional disorder of molecules, making it applicable to describe some aspects of plastic phases. The theory makes the assumption that a molecule in a crystal at a certain temperature may occupy two possible kinds of sites, α and β sites:

- *α sites* are normal sites, which would be all fully occupied in the ordered low-temperature crystal at T = 0K
- *β sites* are the interstitial "abnormal" sites, that are also present in the system.

Between molecules of neighbouring α and β sites, there is a *repulsive* energy, *W*. Moreover each site the molecule can have two different orientations in this model, symbolised by the number "1" and "2". A second *repulsive* energy, *W'*, is associated with pairs of neighbouring molecules in sites α_1 and α_2, within α position. Hence, the molecules can occupy in total four different *site-orientations*: α_1, α_2, β_1, β_2. At 0K, in the equilibrium configuration of the low-temperature order crystal, all molecules would be iso-orientated equivalently in a α_1 site.

The key assumption in the Pople and Karasz model is that the repulsive energy *W*, (which relates to positional disorder) does not depend on the orientations of the molecules, implying that if in a lattice with majority of α_1 configurations, a molecule is turned to a α_2 orientation, then a nearby β site is favoured in lowering the total repulsive energy, by the fact that is not subjected to *W'* repulsive term, which is present only between two

different orientations of identical sites. The partition function, Z, of the system can be evaluated by an extension of the Bragg-William's model for order-disorder phase transitions in alloys[87,88]. Through the partition function is possible to calculate the increase in entropy due to the rotational-positional disorder in the plastic phase, by using the Boltzman's equation in [2.6], but in the form of:

$$S = Nk_B \cdot lnZ \quad , \qquad [2.7]$$

where N is the number of molecules in the system and k_B is the Boltzman's constant.

Pople and Karasz define two parameters Q_p, and S_p, with the first representing the fraction of molecules on α sites, and the second the fraction of molecules in orientation "1" sites. Q_p, and S_p parameters represent respectively the degree of positional order of the system and the degree of orientational order in the system. The model demonstrates that both the positional and orientational order vanish (amorphous liquid) when:

$$S_p = Q_p = \frac{1}{2} \quad . \qquad [2.8]$$

If the condition in equation [2.8] is present, the system is in complete liquid state.

On the other hand, a configuration in which the system with orientational disorder, but positional order appears when:

$$S_p = \frac{1}{2} \; ; \; Q_p > \frac{1}{2} \quad , \qquad [2.9]$$

and the system results in a plastic phase.

Pople and Karasz also define another important quantity, ν, which is the ratio between the energy needed for a reorientation of a molecule, and the energy needed to transfer it into an interstitial site:

$$\nu = \frac{n'W'}{nW} \quad , \qquad [2.10]$$

where n is the number of β sites next to each α site and n' is the number of α sites next to each α site. It has been found that for sufficiently low values of ν, the condition in [2.9], for the presence of a plastic phase, is satisfied at a temperature below the melting point. For large values of ν, the system undergoes a simple 1st order phase transition, at which positional and orientational order disappear together and the system becomes liquid in nature. The critical value of the quantity, ν_c, has been calculated to be 0.325.

The Pople and Karasz model allows deductions on the entropy of melting, ΔS_m, and the relative volume change on melting, $\Delta V/V_s$ (V_s is the volume of the plastic phase close to the melting point). In practice Pople and Karasz's model predicts a decrease in entropy of

fusion, ΔS_m, with a decrease of the relative volume; this is in agreement with experimental data[43].

This model is however limited since it mainly considers only two possible orientations "1" and "2" for each molecule. Moreover, an extension of theory was further considered in order to include the effects of the pressure:

- Under pressure the *ordered crystal – plastic crystal* and the *plastic crystal – liquid* phase transitions can be both of the 2nd order.
- An *ordered crystal – plastic crystal* phase transition, with varying temperature, can occur at high pressures, even if it is not observed at standard pressures.

An interesting study, which validates the model, is represented by the order-disorder plastic crystal phase transition in tetradeutero-methane (CD_4)[89]. This compound presents two orientationally disordered phases I and II and a third ordered phase III. The molecules of CD_4 in the orientationally disordered phase I lie on a face centred cubic lattice and are treated as rigid bodies. Phase II is of particular interest, as one molecule in four rotates quite freely, while surrounded by a shell of oscillating nearest neighbour molecules. The ordered phase III is predicted to have tetragonal symmetry, with all molecules oscillating about equivalent orientations. The theory predicts that ordered phase III – partially disordered phase II is a 1st order transition and phase II – phase I is a 2nd order transition.

Paraffins are interesting systems as plastic phases are very common. A typical characteristic of these systems is the presence of a continuous transition with increasing temperature, of 2nd order type, to a highly packed hexagonal phase, which has typically "liquid crystalline" behaviour[90].

2.4.3 "Glassy" Crystals: "quenching" of plastic crystals

Another interesting property, exhibited by several plastic crystals is their high sensitivity to the kinetic factors involved in the transformation plastic crystal-ordered low temperature crystal, during the cooling of the disordered plastic phase. Indeed, if the latter one is cooled rapidly (quenched), the system does not return to the ordered solid phase, but remains in a "frozen" disordered phase, with the symmetry of the plastic phase preserved[91]. Such structures are generally referred to as "glassy" crystals.

The main difference between glassy phases and traditional amorphous glasses is that the former preserve translational symmetry, whilst the latter not. In fact, the "glassy" crystal preserves hexagonal or cubic symmetry, resulting in rendering the molecules frozen in

different orientations with respect to each other. Nevertheless, local structures appear to be present, which are undergoing an ordering process. These local regions in the material may be present, in which some molecules appeared orientationally ordered and hence have a local symmetry. In the cyanoadamantane family, the "glassy" phase shows local structures with tetragonal symmetry[92]. "Glassy" phases are all characterised by a high fragility[53].

2.4.4 Criterion for predicting occurrence of plastic phases in molecular crystals

A very useful criterion to establish whether a specific compound is likely to show the presence of a plastic phase has been defined on the bases of the volume of free rotational molecules by Postel and Riess[93]. This method differs from that based on molecular shape and its necessary "globular" nature, which has been previously discussed. After a statistical analysis about compounds with "globular" molecular shape, which are able or unable to crystallise in a plastic phase, Postel and Riess noted the determining factor to be related to the degree of molecular interlocking in the low-temperature phase. When interlocking is strong, the formation of a plastic phase by increasing the temperature is prevented, although the molecule may have a "globular" shape. The ratio, R_c, is the key parameter to predict the existence of a plastic phase for a specific compound, which is calculated as follows:

$$R_c = \frac{d_m}{D_M} \quad . \qquad [2.11]$$

In the equation [2.11] d_m is the minimum distance between two molecular centres with mass m in the low-temperature ordered crystal, and D_M is the maximum diameter of the sphere that will be circumscribed by one virtually freely rotating molecule. For values in which $R_c > 0.81$Å, the compound is likely to show an orientational order-disorder plastic phase transition, whereas for values of $R_c < 0.81$Å, the compound is not expected to have a high-temperature plastic phase. The critical value, R_c, does not therefore depend on the temperature of the system.

2.5 Computational Simulation of Plastic Crystals

The majority of investigations of mesophases and in particular of plastic crystals have been performed experimentally. Nevertheless, computer simulations have been attended thanks to the continuous increase in computer performance.

In particular Monte Carlo (MC) and Molecular Dynamics (MD) calculations have been employed in recent years to simulate a large variety of phenomena for both equilibrium and non-equilibrium solid-state system with relative success.

2.5.1 Monte Carlo (MC) simulation

There has been extensive use of MC simulations[94] in the study of phase transitions in recent decades. These include the study of fluids[95], polymers[96] and high-temperature superconductors[97]. MC simulations have been applied also in the study of phase transitions of different molecular crystals, specifically by exploring the nature of a particular system as function of varying temperatures. In particular, isothermal-isobaric MC has been used to investigate the orientational order-disorder phase transformation of adantamane[98] and cubane[47]. In both cases MC has been successful in reproducing the experimental temperature dependent phase transition. It is interesting to note that in the study of the cubane plastic phase a deeper analysis of the nature of the rotational motion of the molecules has been performed, which has yielded a good description of the disordered rotational motion of the molecules in the disordered plastic phase. A further interesting example concerns the investigation of the "glassy" phase of the cyanoadantamane solid state[99]. The computational study allowed a detailed investigation of the nature of the system in three different forms: a cubic plastic phase, a "glassy" orthorhombic phase and a low-temperature monoclinic crystalline phase and provided valuable information about the relative thermodynamic stability of the phases with respect to each other and the role played by fluctuations in the analysis of the formation of tetragonal ordered local structures in the "glassy" phase. The study, however, was limited in scope, as it did not attempt to reproduce the entire phase transition with temperature.

2.5.2 Molecular Dynamics (MD) simulation (see section 3.5)

Molecular Dynamics (MD) simulations are another powerful method in modelling plastic phases[100]. The advantage of this technique with respect to MC is in its intrinsic capability of following the evolution of the system in time, as function of temperature. MD can provide detailed information about the positions, velocities and forces of all atoms present in a given system. This is because MD simulations are based on evaluating the numerical solution of the Newton equations for an ensemble of particles. A more detailed survey about the MD method is treated later in section 3.5.

MD simulations are unfortunately computationally expensive and only allow us to run simulations corresponding to events of very short times (up to the order of microseconds). This severe limitation prevents its use in the study of certain phase transitions, which are very complex (reconstructive) and require the system to overcome high energy barriers, due to bond-breaking and reforming, nucleation, and growth. This limit is strictly related to the *non*-ergodicity of the MD simulated system: the MD simulation cannot perform a full and complete probing of the configurational space of a studied system. Such problems have also been encountered in the simulation of plastic crystals, especially when the rotational dynamics of the molecule is very much slower with respect to the time of the simulation[99]. Long times of simulation (order of nanoseconds) are therefore required for a better description of the rotational phases.

A second, but not secondary problem of the MD and MC techniques, arises from the presence of periodic boundary conditions. In particular, for systems with limited size, periodic boundary conditions can introduce artificial long-range correlations among particles, which can make local fluctuations appear more stable than they effectively are[99]. On the other hand, however, the dimensions of a system strongly affect the computational cost of the simulation. MD simulations have been applied to the study of the plastic phases of a range of compounds, in order to characterise the orientational dynamics of the molecules in the high-temperature plastic phase: after the experimental investigation and characterisation of the plastic phase of succinonitrile[101], Cardini *et al.*[102] performed an MD simulation of the plastic phase. The study gave a deeper insight into the orientational disordered phase, in particular concerning the different conformational molecular changes of the molecule. This serves as an example, in which computational research can go hand-in-hand with experimental work. Ultimately, MD can be a very useful tool to calculate the density of states (D.o.S.) of normal modes associated with molecular and crystal vibrations[103]. These can be obtained from the Fourier transform of a power spectrum of the velocity auto-correlation function, as in the case of the plastic phase of succinonitrile[102] and cyanoadantamane[91] in their plastic phases.

Two other MD studies of the orientationally disordered plastic phase of specific compounds are reported in order to underline the importance of such computational method in the study of plastic phases. The first MD study concerns the investigation of the nature of the plastic phase of sulphur hexafluoride[104] (SF_6). This study revealed that the SF_6 molecules lie on octahedral sites with S-F bonds preferentially aligned, on average, along the axes of the bcc lattice. Moreover, the investigation clarifies the origin

of the disorder in the plastic phase, which was found to arise from two competing interactions acting on each molecule:

a) the interaction between two nearest neighbours in the bcc lattice, which favours the orientational order.
b) the interaction between two next–nearest neighbours (two molecules at adjacent corners of the cubic cell), which is strongly repulsive for this relative orientations.

These two competing interactions generate a frustration effect in the structure of the plastic phase, which, upon cooling, leads to two successive low-temperature phases, representing static and dynamic compromises between the two competing forces.

The second MD study involves the characterisation of the orientationally disordered nature of the plastic phase of *n*-butane[105]. The study revealed the presence of an ordering of the unique axis of the disorderly oriented molecules along a common plane. The study of the rotational dynamics is performed by the refinement of the orientational distribution function (ODF: see its discussion in section *7.3.7.b*), calculated as a linear combination of four angular gaussian distribution functions. A similar study of the ODF of potassium nitrate (KNO_3) disordered phases has been reported, by using a linear combination of spherical harmonic functions[106].

However, it is important to note that none of these methods have attempted to perform an entire simulation of the orientationally order-disorder phase transition of the systems.

2.6 References

(1) Dunitz, J. D. *Pure & Appl. Chem.* **1991**, *63*, 177.

(2) Goryainov, S. V.; Fursenko, B. A.; Belitsky, I. A. *Phys. Chem. Miner.* **1996**, *23*, 297.

(3) Zdanowska-Fraczek, M.; Jakubas, R. *Solid State Commun.* **2005**, *136*, 470.

(4) Ciapala, P.; Jakubas, R.; Bator, G.; Zaleski, J.; Pietraszko, A.; Drozd, M.; Baran, J. *J. Phys.-Condes. Matter* **1997**, *9*, 627.

(5) Dalhus, B.; Gorbitz, C. H. *Acta Crystallogr. Sect. C-Cryst. Struct. Commun.* **1996**, *52*, 1761.

(6) Meyer, R.; Entel, P. *Comput. Mater. Sci.* **1998**, *10*, 10.

(7) Onoda, N.; Matsuo, T.; Suga, H. *Philos. Mag. A-Phys. Condens. Matter Struct. Defect Mech. Prop.* **1988**, *57*, 245.

(8) Mnyukh, Y. *Fundamentals of Solid State Phase Transitions, Ferromagnetism and Ferroelectricty*, 2001.

(9) Mnyukh, Y. V.; Petropav.Nn. *J. Phys. Chem. Solids* **1972**, *33*, 2079.

(10) Cohen, M. D.; Schmidt, G. M. J. *J. Chem. Soc.* **1964**, 1996.

(11) Sukenik, C. N.; Bonapace, J. A. P.; Mandel, N. S.; Lau, P. Y.; Wood, G.; Bergman, R. G. *J. Am. Chem. Soc.* **1977**, *99*, 851.

(12) Iannuzzi, M.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *93*.

(13) Muench, W.; Kreuer, K. D.; Silvestri, W.; Maier, J.; Seifert, G. *Solid State Ion.* **2001**, *145*, 437.

(14) Bailey, M.; Brown, C. J. *Acta Cryst.* **1967**, *22*, 387.

(15) Meier, B. H.; Meyer, R.; Ernst, R. R.; Zolliker, P.; Furrer, A.; Halg, W. *Chem. Phys. Lett.* **1983**, *103*, 169.

(16) Fischer, P.; Zolliker, P.; Meier, B. H.; Ernst, R. R.; Hewat, A. W.; Jorgensen, J. D.; Rotella, F. J. *J. Solid State Chem.* **1986**, *61*, 109.

(17) Meier, B. H.; Ernst, R. R. *J. Solid State Chem.* **1986**, *61*, 126.

(18) Beyer, T.; Price, S. L. *J. Phys. Chem. B* **2000**, *104*, 2647.

(19) Davey, R. J.; Maginn, S. J.; Andrews, S. J.; Black, S. N.; Buckley, A. M.; Cottier, D.; Dempsey, P.; Plowman, R.; Rout, J. E.; Stanley, D. R.; Taylor, A. *J. Chem. Soc.-Faraday Trans.* **1994**, *90*, 1003.

(20) Meier, B. H.; Graf, F.; Ernst, R. R. *J. Chem. Phys.* **1982**, *76*, 767.

(21) Dove, M. T. *Introduction to lattice dynamics*; Cambridge University Press, 1993.

(22) Barnett, S. A.; Broder, C. K.; Shankland, K.; David, W. I. F.; Ibberson, R. M.; Tocher, D. A. *Acta Cryst. Sec. B* **2006**, *62*, 287.

(23) Coppens, P.; Sabine, T. M. *Mol. Cryst.* **1968**, *3*, 507.

(24) Harada, J.; Axe, J. D.; Shirane, G. *Phys Rev. B* **1971**, *4*, 155.

(25) Shirane, G.; Axe, J. D.; Harada, J.; Remeika, J. P. *Phys. Rev. B* **1970**, *2*, 155.

(26) Cowley, R. A.; Buyers, W. J. L.; Dolling, G. *Solid State Commun.* **1969**, *7*, 181.

(27) Dove, M. T.; Heilmann, I. U.; Kjems, J. K.; Kurittu, J.; Pawley, G. S. *Phys. Status Solidi B-Basic Res.* **1983**, *120*, 173.

(28) Mackenzie, G. A.; Pawley, G. S. *J. Phys. C- Solid Science Phys.* **1979**, *12*, 2717.

(29) Dietrich, O. W.; Mackenzie, G. A.; Pawley, G. S. *J. Phys. C-Solid State Phys.* **1975**, *8*, L98.

(30) Blinc, R.; Levanyuk, A. P. *Incommensurate Phases in Dielectrics volume 2: Materials*; Amsterdam: North-Holland, 1986.
(31) Bailey, S. W.; Frank-Kamenetskii, V. A.; Goldsztaub, S.; Kato, A.; Pabst, A.; Schultz, H.; Taylor, H. F. W.; Fleischer, M.; Wilson, A. J. C. *Acta Cryst. Sec. A* **1977**, *33*, 681.
(32) Tuble, C. S.; Anwar, J.; Gale, J. D. *J.Am.Chem.Soc.* **2004**, *126*, 396.
(33) Hull, D. *Introduction to Dislocations*; Pergamon Press:: Oxford, 1975.
(34) Cruceanu, E.; Niculescu, D.; Ivanciu, O. *J. Mater. Sc. Lett.* **1974**, *9*, 1391.
(35) Mnyukh, Y. V.; Panfilova, N. A.; Petropavlov, N. N.; Uchvatovski, N. S. *J. Phys. Chem. Solids* **1975**, *36*, 127.
(36) Bhattacharya, K.; Conti, S.; Zanzotto, G.; Zimmer, J. *Nature* **2004**, *428*, 55.
(37) Soltzber.Lj; Ash, B. M.; McKay, P. C. *Mol. Cryst. Liq. Cryst.* **1973**, *21*, 283.
(38) Kaneko, F.; Kobayashi, M.; Sato, K.; Suzuki, M. *J. Phys. Chem. B* **1997**, *101*, 285.
(39) Reinitzer, F. *Monaths. Chem.* **1888**, *9*, 421.
(40) Timmermans, J. *J. Phys. Chem. Solids* **1961**, *18*, 1.
(41) Timmermanns, J. *Bull.Soc. Chem. Belg.* **1935**, *44*, 17.
(42) Gray, G. W.; Winsor, P. A. *Liquid Crystals and Plastic Crystals*; Ellis Horwood:: Chichester, 1974; Vol. I.
(43) Staveley, L. A. *Annu. Rev. Phys. Chem.* **1962**, *13*, 351.
(44) Powles, J. G.; Williams, D. E.; Smyth, C. P. *J. Chem. Phys.* **1953**, *21*, 136.
(45) Mack, J. W.; Torchia, D. A. *J. Phys. Chem.* **1991**, *95*, 4207.
(46) Sherwood, J. N. *The Plastically Crystalline State* Chichester, 1979.
(47) Murugan, N. A. *J. Phys. Chem. B* **2005**, *109*, 23955.
(48) Post, B. *Acta Cryst.* **1959**, *12*, 349.
(49) Dunning, W. J. *J. Phys. Chem. Solids* **1961**, *18*, 21.
(50) Post, B.; Schwartz, R. S.; Fankuchen, I. *J. Am. Chem. Soc.* **1951**, *73*, 5113.
(51) Brand, R.; Lunkenheimer, P.; Loidl, A. *J.Chem. Phys.* **2002**, *116*, 10386.
(52) Smith, C. P. *Dielectric Behaviour and Structure*; McGraw-Hill: New York, 1955.
(53) Pardo, L. C.; Lunkenheimer, P.; Loidl, A. *J. Chem. Phys.* **2006**, *124*.
(54) Smith, G. W. "International Science and Technology," January 1967.
(55) Powles, J. G.; Gutowsky, H. S. *J. Chem. Phys.* **1953**, *21*, 1695.
(56) Smith, G. W. *J. Chem. Phys.* **1965**, *43*, 4325.
(57) Aston, J. G. *J. Phys. Chem. Solids* **1961**, *18*, 62.
(58) Lopez, D. O.; Barrio, M.; Tamarit, J. L.; Negrier, P.; Haget, Y. *Mol. Cryst. Liq. Cryst. Sci. Technol. Sect. A-Mol. Cryst. Liq. Cryst.* **1995**, *268*, 129.
(59) Barrio, M.; Lopez, D. O.; Tamarit, J. L.; Haget, Y. *Mater. Res. Bull.* **1995**, *30*, 659.
(60) Barrio, M.; Font, J.; Lopez, D. O.; Muntasell, J.; Tamarit, J. L.; Negrier, P.; Chanh, N. B.; Haget, Y. *J. Phys. Chem. Solids* **1993**, *54*, 171.
(61) Suga, H.; Sugisaki, M.; Seki, S. *Mol. Cryst.* **1966**, *1*, 377.
(62) Farid, M. M.; Khudhair, A. M.; Razack, S. A. K.; Al-Hallaj, S. *Energy Conv. Manag.* **2004**, *45*, 1597.
(63) Khudhair, A. M.; Farid, M. M. *Energy Conv. Manag.* **2004**, *45*, 263.
(64) Chandra, D.; Chien, W. M.; Gandikotta, V.; Lindle, D. W. *Z. Phys. Chemie-Int. J. Res. Phys. Chem. Chem. Phys.* **2002**, *216*, 1433.
(65) Font, J.; Muntasell, J.; Cesari, E.; Pons, J. *J. Mater. Res.* **1997**, *12*, 3254.

(66) Barrio, M.; Font, J.; Lopez, D. O.; Muntasell, J.; Tamarit, J. L.; Haget, Y. *J. Appl. Crystallogr.* **1994**, *27*, 527.
(67) Muntasell, J.; Barrio, M.; Font, J.; Lopez, D. O.; Tamarit, J. L. I.; Diarte, M. A. C.; Guion, J.; Teisseire, M.; Chanh, N. B.; Haget, Y. *J. Therm. Anal.* **1991**, *37*, 2395.
(68) Long, S.; MacFarlane, D. R.; Forsyth, M. *Solid State Ion.* **2003**, *161*, 105.
(69) Cooper, E. I.; Angell, C. A. *Solid State Ion.* **1986**, *18-9*, 570.
(70) Alarco, P. J.; Abu-Lebdeh, Y.; Abouimrane, A.; Armand, M. *Nat. Mater.* **2004**, *3*, 476.
(71) Abu-Lebdeh, Y.; Alarco, P. J.; Armand, M. *J. New Mat.Electrochem. Syst.* **2004**, *7*, 29.
(72) Abu-Lebdeh, Y.; Alarco, P. J.; Armand, M. *Angew. Chem.-Int. Edit.* **2003**, *42*, 4499.
(73) Alarco, P. J.; Abu-Lebdeh, Y.; Armand, M. *Solid State Ion.* **2004**, *175*, 717.
(74) Weinstock, B. *J. Phys. Chem. Solids* **1961**, *18*, 86.
(75) Marshall, J. G.; Staveley, L. A. K.; Hart, K. R. *Trans. Faraday Soc.* **1956**, *52*, 19.
(76) Staveley, L. A. K. *J. Phys. Chem. Solids* **1961**, *18*, 46.
(77) Smyth, C. P. *J. Phys. Chem. Solids* **1961**, *18*, 40.
(78) Cardini, G.; Ricci, M.; Righini, R.; Califano, S. *Chem. Phys.* **1994**, *189*, 17.
(79) Cardini, G. *Chem. Phys.* **1995**, *193*, 101.
(80) Andrew, E. R. *J. Phys. Chem. Solids* **1961**, *18*, 9.
(81) Yukitoshi, T.; Suga, H.; Seki, S.; Itoh, J. *J. Phys. Soc. Jpn.* **1957**, *12*, 506.
(82) Eldrup, M.; Lightbody, D.; Sherwood, J. N. *Faraday Discuss.* **1980**, 175.
(83) Guthrie, G. B.; McCullough, J. P. *J. Phys. Chem. Solids* **1961**, *18*, 53.
(84) Landau, L. D.; Lifshitz, E. M. *Statistical Physics: Course of Theoretical Physics*; Pergamon Press, 1969; Vol. 5.
(85) Karasz, F. E.; Pople, J. A. *J. Phys. Chem. Solids* **1961**, *20*, 294.
(86) Pople, J. A.; Karasz, F. E. *J. Phys. Chem. Solids* **1961**, *18*, 28.
(87) Bragg, W. L.; Williams, E. J. *Proc. Royal. Soc. Lond. Series A* **1935**, *151*, 699.
(88) Bragg, W. L.; Williams, E. J. *Proc. R. Soc. Lond., Series A* **1935**, *151*, 540.
(89) James, H. M.; Keenan, T. A. *J. Chem. Phys.* **1959**, *31*, 12.
(90) Ungar, G. *Macrom.* **1986**, *19*, 1317.
(91) Descamps, M.; Willart, J. F.; Kuchta, B.; Affouard, F. *J. Non-Cryst. Solids* **1998**, *235*, 559.
(92) Luty, T.; Rohleder, K.; Lefebvre, J.; Descamps, M. *Phys. Rev. B* **2000**, *62*, 8835.
(93) Postel, M.; Riess, J. G. *J. Phys. Chem.* **1977**, *81*, 2634.
(94) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquid*; Oxford Science: Oxford, 1987.
(95) Schoen, M.; Diestler, D. J.; Cushman, J. H. *Mol. Phys.* **1993**, *78*, 1097.
(96) Lai, P. Y. *Macromol. Theory Simul.* **1999**, *8*, 382.
(97) Gracheva, M. E.; Katargin, M. V.; Kashurnikov, V. A.; Rudnev, I. A. *Low Temp. Phys.* **1997**, *23*, 863.
(98) Yashonath, S.; Rao, C. N. R. *J. Phys. Chem.* **1986**, *90*, 2552.
(99) Kuchta, B.; Descamps, M.; Affouard, F. *J. Chem. Phys.* **1998**, *109*, 6753.

(100) Haile, J. M. *Molecular Dynamics Simulation:Elementary Methods*; Wiley & Sons: United States, 1992.

(101) Derollez, P.; Lefebvre, J.; Descamps, M.; Press, W.; Fontaine, H. *J. Phys.-Condes. Matter* **1990**, *2*, 6893.

(102) Cardini, G.; Righini, R.; Califano, S. *J. Chem. Phys.* **1991**, *95*, 679.

(103) Winkler, B.; Dove, M. T. *Phys. Chem. Minerals* **1992**, *18*, 407.

(104) Dove, M. T.; Pawley, G. S. *J. Phys. C-Solid State Phys.* **1984**, *17*, 6581.

(105) Pawley, G. S.; Refson, K. *Acta Cryst. Sec. A* **1987**, *43*, 727.

(106) Smith, W.; Clarke, J. H. R. *J.Chem.Phys.* **1989**, *90*, 6610.

Chapter 3

METHODS OF SIMULATION AND PROGRAMS

3.1 Introduction

The prediction of organic molecular crystal structure is made through a systematic procedure, which has been already schematised in the flow-chart in figure 1.7 of chapter 1. Different simulation packages are used in the different steps of the search procedure and in the present chapter the methodologies and the packages used in the search are discussed. First there is a brief description of the *ab initio* method used in the optimisation of the molecular geometry and the calculation of electronic structure, and then attention is focused on the search method, lattice energy minimisation algorithm, and molecular dynamics (MD) algorithms. Furthermore a detailed explanation of the methods used to calculate X-ray patterns from the MD configurations is presented. Finally, in the appendixes at the end of the chapter, we give a brief explanation of the Ewald summation treatment in modelling the electrostatic contribution, the calculation of second derivative properties of predicted crystal structures (phonon frequencies and elastic constants) and the calculation of the volume growth rate.

3.2 *Ab-Initio* Molecular Calculation: GAUSSIAN 98

This program performs quantum mechanical calculations on molecules, using both Hartree-Fock (HF) and Density Functional Theory (DFT) methods. In this thesis the HF equations are used and the electron correlation effects are treated at the MP2 level of theory [1]. The same method is used to calculate the molecular electronic structure, with different methods, such as Mulliken population analysis[2] and methods based on fitting the electronic charges to the molecular electrostatic potential (MEP[3]). In these studies we follow the latter method, using the CHelpG algorithm[4], which uses the following method to fit the atomic charges to the values of the electrostatic potential in the region around the molecule: it defines a cube of points spaced 0.3-0.8 Å apart, containing the molecule and including 2.8 Å of headspace on all sides of the cube. All points, which fall out of this maximum value, are eliminated from the fitting procedure. All points, which fall inside a predefined van der Waals (vdW), radius of any of the nuclei, are also discarded.

The remaining points form a homogeneous layer around the molecule, where all regions outside the vdW radius and within the defined interaction distance are equally represented. The total number of selected points is ~ 5000 points[4] and at each of the sample points the molecular electrostatic potential is calculated analytically, from the *ab initio* wave function. These data are then used as input for a Lagrange multiplier least-squares routine, which fits the exact molecular charge, but not the dipole moment.

The basis set is a crucial factor in *ab initio* calculation as it essentially determines the form and expression of the wave function of the system, as its quality has a large influence on the reliability of the results of the calculation. The wave function of the system is represented by the Slater determinant[1], which is a sum of products of spin orbitals, which represent the wave function of the individual electrons in the system. These spin orbitals are given by the product of an electron spin function and a spatial function of atomic orbitals (AO), with the latter represented by a combination of a specific set of functions, forming an orthogonal basis set: $\{\phi_i, i = 1...n\}$. Commonly Gaussian functions are used to represent the basis functions, with one or more Gaussians being used to describe each of the AOs, i.e.:

$$g_{1s}(\alpha, r) = \left(\frac{2\alpha}{\pi}\right)^{\frac{3}{4}} \cdot e^{-\alpha r^2} \quad , \qquad [3.1]$$

$$g_{2p_x}(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3}\right) \cdot r \cdot e^{-\alpha r^2} \quad . \qquad [3.2]$$

As more Gaussian functions are added to the basis set the computer time needed to calculate the integrals in the Fock operator[1] increases rapidly[1]. To avoid large basis sets, Contracted Gaussian Function (CGF) are commonly used (figure 3.1). They are linear combinations of primitive gaussian functions (PGF) as reported in formula [3.3], with fixed coefficients, which do not need to be optimised during the variational procedure. These new functions make the description of the electrons in the core and in the valence regions of the molecule more accurate and efficient from a computational point of view.

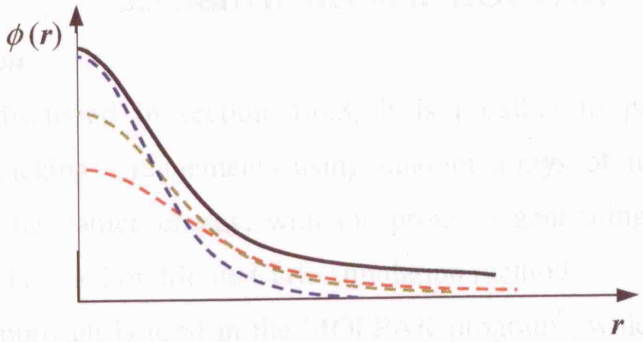


Fig. 3.1: Example of CGF basis functions.

$$\phi^{CGF}{}_{\mu}(r) = \sum_{i=1}^{n} d_{i\mu} \cdot g_{\mu}(\alpha_{i\mu}, r) \quad , \qquad [3.3]$$

where μ=1s, 2s, 2p, 3d, 3f .....and d are the fixed coefficients.

The basis set used in our calculations of the molecular electronic structure and the molecule optimisation is 6-31G**, which is a Split-Valence basis set, which is a set of CGF basis functions, consisting of:

- 1 CGF function, combination of 6 gaussian functions for the electrons of the "core" shell of the atom.
- 2 functions for the electrons in the "valence" shell, where one of the functions is a combination of 3 gaussian functions and the second is a primitive function.

The two stars indicate that are also polarized functions:

- 6 gaussian functions for the d orbitals (α = 0.8 for the atoms C, N, O, F)
- one primitive gaussian function of p type (α = 1.1) for each H atom.

The optimisation of the molecular geometry, which has a large influence in determining the properties of a system, yields the most stable geometry of the isolated molecule, which is the minimum in energy. The optimisation is performed using the following procedure: starting from an initial set of atomic positions, which describe the molecular conformation, a SCF calculation is performed to calculate the energy and its first derivative (or energy gradient) with respect to the atomic coordinates. If the latter is zero, the energy of the system is at a minimum and nothing further needs to be done, while, if it is not zero, the atomic positions will be varied based on the size and direction of the energy gradient, and a new SCF calculation will be performed at a geometry with a lower energy, until the system converges to a minimum energy.

3.3 Search Method: MOLPAK

3.3.1 Introduction

As previously discussed in section 1.6.3, it is possible to predict different crystal structures and packing arrangements using random arrays of molecules following by minimisation of the lattice energy, with the program generating the random arrays of molecule, which is based on Monte Carlo simulation method[5].

An alternative approach is used in the MOLPAK program[6], which predicts the possible crystal structures for a molecule, following the criteria of minimum volume, or of maximum density. It considers the molecules as having fixed and rigid conformations.

The unit cell is the basic unit in a crystal and it is sufficient to know its size and shape to describe completely the crystal structure. The unit cell has an internal symmetry. Often it is useful to use a smaller part of the unit cell to describe the crystal structure, i.e. the *asymmetric unit*, which is the smallest part of a crystal structure, to which symmetry operators can be applied to generate the entire crystal. These symmetry operators relate each asymmetric unit to the others contained in the unit cell.

Hence, a crystal structure can be described specifying:

- unit cell's dimensions
- space group
- structure of the asymmetric unit.

We also have a second way to describe a crystal structure:

- place the centre of mass of a molecule at the origin of an orthogonal coordinate system
- specify the molecular environment surrounding this central molecule, that is the *coordination sphere* of the central molecule. The orientations of these molecules relate to the orientation of the central molecule by the symmetry operators.

When there is one molecule per asymmetric unit cell, the following equation is always valid for a crystal:

$$N^{\circ}_{molecule}\ per\ unit\ cell = N^{\circ}_{symmetry\ operators} \quad . \qquad [3.4]$$

MOLPAK uses this second way to form all the possible crystal structures of a certain molecule, treating:

- One molecule per asymmetric unit.

- Symmetry operators: identity(1), inversion centre(-1), two-fold screw axes(2_1), glide plane/mirror plane(*m*).

All the unit cells operations involve

- Change of orientation around a fixed centroid molecule
- Translation of the centroid to a new position molecule

The coordination sphere is composed of molecules, which are in close contact with the central molecule and are involved in a van der Waals contact with centre of mass of this molecule.

Table 3.1 reports the typical van der Waals (vdW) radii for the most common atoms in organic molecules.

Element	*Van der Waals Radius /Å*
N	1.55
C	1.70
H	1.10
O	1.52
F	1.50

Table 3.1: Van der Waals radii of the most typical elements in organic solid state.

A molecular packing analysis[6] of the coordination spheres in the crystal structure of energetic compounds has focused onto the minimum distance between the vdW spheres (typical radii are reported in table 3.1) of the central molecule and the vdW spheres of the surrounding molecules, as function of the number of coordinating molecules (n_c). The analysis established a sharp increase of the minimum distance occurring for $n_c > 14$ and proposed the latter as the largest number of coordinating molecules to be considered in order to maintain close packed distribution.

The composition and the motif of the sphere is characteristic of each different space group. The crystal structures can be obtained by building the pattern of these coordination spheres around the central molecule.

3.3.2 Packing Procedure

The packing procedure used by this program in building all the structures consists of three steps.

1) A central molecule is placed at the origin of a Cartesian reference system (*x, y, z*).

Then a new molecule is moved towards the central molecule along the x axis, until some threshold value in the *repulsion energy* is satisfied.

If $E_{rep} < E^{(1)}{}_{Threshold}$	If $E_{rep} > E^{(1)}{}_{Threshold}$
The second molecule is moved closer to the central molecule in arbitrary equal steps until $E_{rep} > E^{(1)}{}_{Threshold}$	The second molecule is moved away in arbitrary equal steps until $E_{rep} < E^{(1)}{}_{Threshold}$

In this way the *reference position* is obtained for the second molecule. Then the arbitrary step, previously chosen, is halved so that the second molecule moved closer to the central molecule and a new evaluation of the repulsive energy is performed.

If $E_{rep} < E^{(1)}{}_{Threshold}$ the new position of the second molecule becomes the new *reference position*. This procedure continues until the step size reaches a certain fraction ($\sim \frac{1}{128}$) of the initial step. Once the position of the second molecule is established, a line of central molecules along the x axis is automatically created (see figure 3.2), with a fixed unit translation.

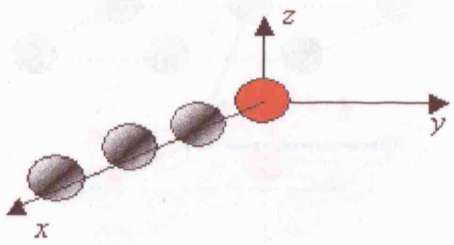


Fig.3.2: Schematic illustration of the 1th step of MOLPAK in building the coordination sphere of a central molecule.

2) A 2D-grid of molecules is established by moving a line of molecules toward the central molecule along the y axis, until a certain $E^{(2)}{}_{Threshold}$ is reached. Considering also the space group symmetry, the smallest unit cell in 2D is defined in this step, so that the ***a***, ***b*** cell parameters and the angle γ between them are defined (see figure 3.3).

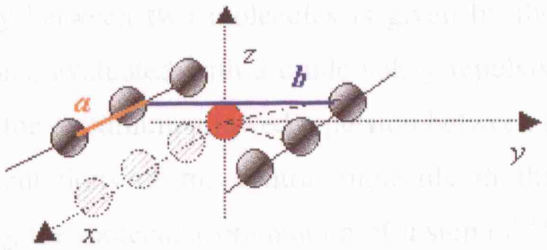


Fig. 3.3: Schematic illustration of 2th step of MOLPAK in building the coordination sphere of a central molecule.

Because the central molecule is usually in a close contact with two molecules of the adjacent line, this relationship is valid for the repulsion energy threshold:

$$E^{(2)}_{Threshold} = 2\, E^{(1)}_{Threshold} \qquad [3.5]$$

3) Finally a 2D- grid (like that obtained in the step 2) is moved parallel to the central molecule's grid (the central grid), along the *z* axis, in order to find a minimum distance consistent with a repulsion energy maximum value $E^{(3)}_{Threshold}$ (see figure 3.4).

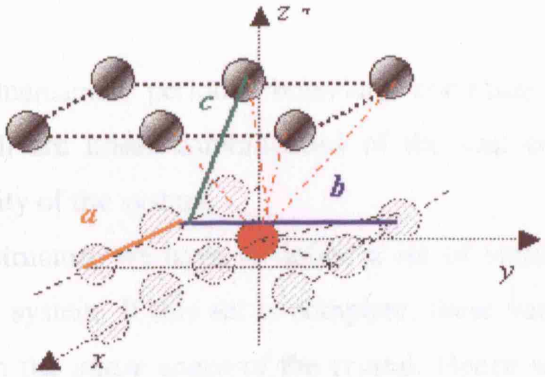


Fig. 3.4: Schematic illustration of the 3th and final step of MOLPAK in building the coordination sphere of a central molecule.

Considering the space group symmetry, this step defines the third cell parameter (***c***) and the two angles α and β, respectively between ***a***, ***c*** and ***b***, ***c***, as shown in the figure above. Because the central molecule is usually in close contact with four molecules, belonging to the upper grid, this relationship is valid for the repulsion energy threshold:

$$E^{(3)}_{Threshold} = 4\, E^{(1)}_{Threshold} \qquad [3.6]$$

The repulsion energy between two molecules is given by the sum of all the pair-wise atom-atom interactions, evaluated with a crude solely repulsive potential. The threshold used in the step 1) is the maximum allowed repulsion between two molecules.

All the possible orientations of the central molecule in the coordination sphere are considered by varying the molecular orientation of a step of 15 degrees around its centre of mass. For each new orientation the 3-step packing operation described above is repeated in order to produce a 3D-map of minimum unit cell volume as function of the central molecule's orientation. For each space group 200 possible coordination spheres are generated from this procedure.

MOLPAK uses rigid-body molecular structures as search probes in the construction of a packing pattern.

3.4 Lattice Energy Minimisation: DMAREL

This program minimises the lattice energy of the unit cell of crystal structures. In doing so it is not constrained within the particular space group symmetry. Nevertheless, the program has the option to relax the symmetry of the system to P1. However, it treats the molecules inside the unit cell as rigid entities, to reduce the number of independent variables.

The program uses 3-dimensional periodic boundary condition - there are specific translation vectors, which are linear combinations of the unit cell lattice parameters, which define the periodicity of the system.

To optimise the crystal structure we have to define a set of variables, representing the degree of freedom of the system. If this set is complete, these variables can be used as orthogonal axes and form the *phase space* of the crystal. Hence we can map the lattice energy as function of this set of variables, so that we have an energy surface on this *phase space*. A minimum in this energy surface represents a possible stable crystal structure.

3.4.1 Set of Variables

The orthogonal set of variables representing the *phase space* of the crystal is given by:

- *Positional Coordinates*: we consider the components of the centre of mass for each molecule $\boldsymbol{r}^{(CM)}_i$, $i = 1, \ldots N$, where N is the number of molecules in the unit cell.
- *Orientation Coordinates*: these are used to describe the molecular orientation. For each molecule we consider three components of a vector $\boldsymbol{\Theta}_i = (\Theta_1, \Theta_2, \Theta_3)$,

which passes through the centre of mass of the molecule. This vector describes an axis of rotation for the molecule. The changes in the orientation of the molecule are described using a rotation matrix (Q), operating on all the vectors s, defining the molecule. Called $s_0 = (s_{01}, s_{02}, s_{03})$ the original vector, the new vector (after the rotation) is:

$$s = Qs_0 , \qquad [3.7]$$

where Q is the rotation matrix:

$$Q = \begin{pmatrix} e_0^2 + e_1^2 - e_2^2 - e_3^2 & 2(e_1e_2 - e_0e_3) & 2(e_1e_3 + e_0e_2) \\ 2(e_1e_2 + e_0e_3) & e_0^2 - e_1^2 + e_2^2 - e_3^2 & 2(e_2e_3 - e_0e_1) \\ 2(e_1e_3 - e_0e_2) & 2(e_2e_3 + e_0e_1) & e_0^2 - e_1^2 - e_2^2 + e_3^2 \end{pmatrix} \qquad [3.8]$$

$$e_0 = cos\left(\frac{|\Theta|}{2}\right)$$

$$e_i = \frac{\Theta_i}{|\Theta|} sin\left(\frac{|\Theta|}{2}\right) \qquad i = 1, 2, 3 \qquad [3.9]$$

- <u>*External Coordinates*</u>: these are used to describe the geometry of the unit cell. The changes in the cell dimensions are described by a symmetric strain matrix, which gives the changes in all the three lattice vectors ***a, b, c***. Defining with r_0, the position vector of a molecule in the unit cell, before the action of the strain matrix E on the crystal, a change of the crystal geometric shape is represented by:

$$r = r_0 + Er_0 , \qquad [3.10]$$

where the strain matrix is defined in the following way[7]:

$$E = \begin{pmatrix} E_1 & \frac{1}{2}E_6 & \frac{1}{2}E_5 \\ \frac{1}{2}E_6 & E_2 & \frac{1}{2}E_4 \\ \frac{1}{2}E_5 & \frac{1}{2}E_4 & E_3 \end{pmatrix} . \qquad [3.11]$$

In summary, the set of variables comprises 6 coordinates for each molecule in the unit cell (3 for the positional vector $r^{(CM)}$ and 3 for the orientation vector Θ) and 6 strain matrix elements to describe the changes in the unit cell. Thus for a unit cell of N molecules, the total number of variables is $6N+6$, and each point on an energy surface represents a crystal structure.

3.4.2 Minimisation Procedure

- *Newton-Raphson Procedure*

The search for the minimum in phase space is performed using a modified Newton-Raphson procedure. The lattice energy U is expressed as a second order Taylor expansion in the set of variables.

$$U(x_1+\delta x_1,\ldots x_i+\delta x_i,\ldots x_{(6N+6)}+\delta x_{(6N+6)}) = U(x_1,\ldots x_i,\ldots x_{(6N+6)}) + \sum_{i=1}^{6N+6}\frac{\partial U}{\partial x_i}\delta x_i + \frac{1}{2}\sum_{i=1}^{6N+6}\sum_{j=1}^{6N+6}\delta x_i\frac{\partial^2 U}{\partial x_i \partial x_j}\delta x_j \qquad [3.12]$$

The vector $\delta \boldsymbol{x} = (\delta x_1, \delta x_2, \delta x_3, \ldots \delta x_i, \ldots \delta x_{6N+6})$, represents the infinitesimal displacement along the energy surface of the crystal. At an energy minimum, the following conditions must be satisfied:

$$\left(\frac{\partial U}{\partial x_i}\right) = 0 \ , \qquad [3.13]$$

$$\left(\frac{\partial^2 U}{\partial x_i \partial x_j}\right) > 0 \ , \qquad [3.14]$$

where $i, j = 1,2,3, \ldots 6N+6$.

At equilibrium, the derivative of the lattice energy with respect to any of the coordinates of the displacement $\delta \boldsymbol{x}$ must be zero, which leads to the following set of equations for the displacement leading to the equilibrium configuration:

$$0 = 0 + \frac{\partial U}{\partial x_i} + \frac{1}{2}(W + W^T)_{ij}\delta x_j \qquad i = 1, 2, 3, \ldots .6N+6 \ , \qquad [3.15]$$

where

$$W = \sum_{j=1}^{6N+6}\frac{\partial^2 U}{\partial x_i \partial x_j} \qquad [3.16]$$

is the second derivative matrix (or Hessian) of the energy.

The displacement for a certain position of the *minimum* (p) is given by:

$$\delta \boldsymbol{x}_p = -2\sum_{i=1}^{6N+6}(W + W^T)_{ij}^{-1} \cdot \frac{\partial U}{\partial x_i} \ . \qquad [3.17]$$

The orientation-orientation second derivatives introduce asymmetry in the W matrix, but the matrix $(W+W^T)$ is always symmetric. The calculated displacement $\delta \boldsymbol{x}_p$ remains unique for a given set of variables. If the potential energy surface were perfectly harmonic (i.e. the second order Taylor expansion is valid), $\delta \boldsymbol{x}_p$ would locate the minimum in one step.

In practice, the energy surface will be more complex, but will become increasingly close to harmonic as we approach the minimum.

- *Iterative Process*

On the 6N+6 dimensional phase space, the system is in an initial point on the energy surface and we wish to move to a point of minimum energy (local minimum) on the surface, which we can do through an iterative process, which involves the choice of the direction of the first step of the minimisation, and then estimate the best value for the step. The process is schematically illustrated in figure 3.5:

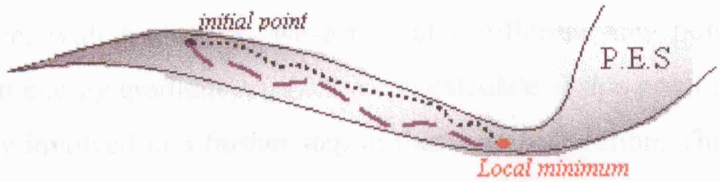


Fig. 3.5: Potential energy surface (PES) in the 3D phase space for a crystal and schematic illustration of the search of the direction path, leading to a local minimum.

In calculating the step direction $\delta \boldsymbol{x}_p$, using equation [3.17], we impose the following constraint:

$$\sum_{p=1}^{3N} \delta \boldsymbol{x}_p = 0 \quad , \qquad [3.18]$$

where the $3N$ degree of freedom in equation [3.18] are the centre of mass position coordinates $\boldsymbol{r}^{(CM)}_i$ (i =1, …N). This implies constraining the net translation of the unit cell to be zero.

As noted above the Taylor expansion of the energy steps is truncated at the second order so that equation [3.17] for the displacement is exact only in the case of a quadratic energy surface. We can always assume that the surface is at least locally quadratic and in this way, starting from an initial point, we can use the vector $\delta \boldsymbol{x}_p$ as search direction thorough an iterative process, to find the nearest local minimum on the energy surface. At each iterative step, the displacement $\delta \boldsymbol{x}_p$, as defined by equation [3.19], is updated by a scaling factor α.

$$\delta \boldsymbol{x}_p^{(\text{new})} = \alpha \, \delta \boldsymbol{x}_p \quad . \qquad [3.19]$$

Hence, the minimisation procedure has to search also for the best value of the scaling factor α. The calculation of the inversion matrix in equation [3.17] is computationally intensive and time consuming.

- *Energy Drop*

It is possible to calculate the *energy drop* (or *energy gradient*) along the search direction, from the gradient vector and the displacement vector, using equation [3.20]:

$$grad_{\ n} = \sum_{i=1}^{m} \left(\frac{\partial U}{\partial x_i} \right) \alpha_n \delta \boldsymbol{x}_p^{(i)}, \qquad [3.20]$$

where n is the n^{th} value of the scaling factor α: i.e., α_1 is the value of the scaling factor after the first update of the displacement; α_2 is the value after the latest move along the search direction, and so on and m is the maximum number of the steps performed to reach the local minimum. If we move (after the first update) along the search direction on the energy surface, with a step $\delta \boldsymbol{x}_p$ we arrive at a different new point on the energy surface, having an energy gradient $grad_2$. We can calculate at this point a new estimate of the drop in energy involved in a further step in that search direction. There are 3 possible trends of the energy gradient after the first step along a certain direction:

a) $\begin{cases} |grad_1| > |grad_2| \\ grad_1 < 0;\ grad_2 < 0 \end{cases}$ This case is illustrated in figure 3.6.

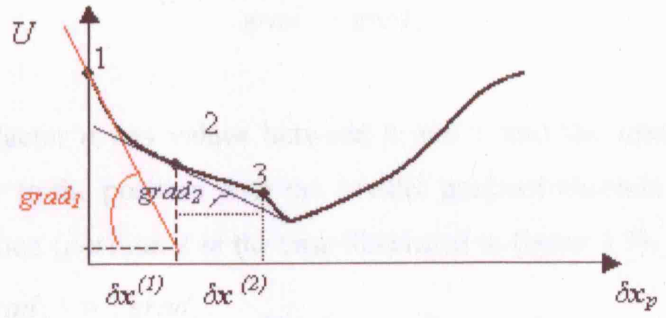


Fig. 3.6: Lattice energy drop along a search direction: case a) $grad_1 < grad_2 < 0$.

In this case the step in the search direction is approaching the minimum value possible along that direction, but it has not reached it. The program tests the point, to see if the energy drop ($grad_1$) is large enough to justify a new search in the chosen direction. The condition is:

$$|grad_2| < \frac{1}{4}|grad_1| \qquad . \qquad [3.21]$$

If this condition is not satisfied, then the system will be moved by a further step.

b) $\begin{cases} grad_1 < 0 \\ grad_2 > 0 \end{cases}$ This case is illustrated in figure 3.7.

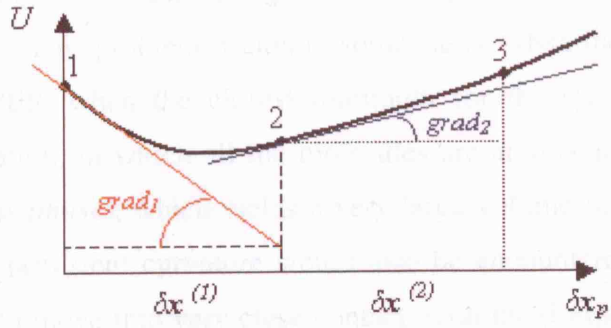


Fig. 3.7: Lattice energy drop along a search direction: case b) $grad_1 < 0$ and $grad_2 > 0$.

In this case the step δx has passed over the minimum, along the search direction. The program performs a linear interpolation, by scaling the initial step by a factor:

$$\alpha = -\frac{grad_1}{grad_2 - grad_1} - 1 \qquad [3.22]$$

The factor α has values between 0 and 1 and the interpolation gives a point closer to the position with the smaller gradient absolute value along the search direction (position 2 in the case illustrated in figure 3.7).

c) $\begin{cases} |grad_2| > |grad_1| \\ grad_1 < 0;\ grad_2 < 0 \end{cases}$ This case is illustrated in figure 3.8.

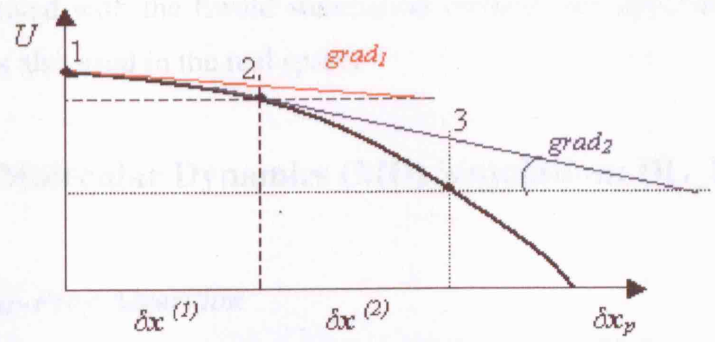


Fig. 3.8: Lattice energy drop along a search direction: case c) $grad_2 < grad_1 < 0$.

In this last case the gradient becomes increasing negative along the search direction and no minimum in the energy is found. The program may continue to find better areas of the potential surface, but if this situation persists, it stops to run, giving the warning signal of the presence of a *persistent negative curvature*. This problem occurs in some cases, when there are unstable regions on the PES, when the closest minimum for the system corresponds to the configuration, in which all the molecules are at very large distance from each other (*gas phase*), which yields a very large volume of the final unit cell. The negative persistent curvature would also be encountered when atom pairs are allowed to move into very close contact, with the *6-exp* potential going to the -∞, (see section 1.5.2).

The interpolated value of α is used to calculate the gradient at the new position, at the further step, so that $grad_2$ value is always reset (it is always the last point to be tested).

The energy and its derivatives are calculated in the limit of the *pairwise* approximation (see chapter 2), which considers atom-atom pair interactions. The crystal energy is given so by the relationship:

$$U = \sum_{i \in L} \sum_{j \in M} U_{ij} \quad , \qquad [3.23]$$

where U_{ij} is the interaction energy of the atomic sites i and j, belonging to different molecules. In the summation between different centres of mass a cut-off of 15Å is used, which has been shown to give a good convergence in the calculation of the lattice energy minimum as function of the cut-off. This summation scheme is applied to all of the multipoles interactions (usually including the interactions up to the order of $\boldsymbol{R}^{-5}$, as previously discussed in section 1.4.2) except for charge-charge, charge-dipole and dipole-dipole interactions. In these last three cases, the summation converges only slowly, so that they are treated with the Ewald summation method (see appendix A), for which a cut-off of 15Å is also used in the real space.

3.5 Molecular Dynamics (MD) Simulation: DL_POLY

3.5.1 Verlet Leap-Frog Algorithm

Energy minimization techniques are useful to obtain the equilibrium state of the system, but this approach has many limitations. For example the effect of temperature is not taken into account and time plays no role at all, so it is not possible to obtain information about

any dynamical processes. The most obvious way of introducing both time and temperature in our treatment of the atoms is by using Newton's equations of the motion for each atom *i*:

$$m_i \frac{d^2 \boldsymbol{r}_i}{dt^2} = -\nabla_i V_{total}(\boldsymbol{r}_1,....,\boldsymbol{r}_N) \ , \qquad i = 1, ...,N \qquad [3.24]$$

where V_{total} is the potential energy of the system. This technique is called molecular dynamics (MD)[8,9].

The connection between the thermodynamic properties and the motion of atoms is not generally straightforward. For example, although the estimate of the temperature is easy, as it is directly related to the atomic velocities, the calculation of the other properties such as free energies is more complex. We use DL_POLY[10] and DL_MULTI[11,12], which is a version of DL_POLY_2, which allows us to treat the electrostatics using a distributed multiple analysis (DMA[13,14]), which has been previously discussed in section 1.4.2.

The algorithm used in solving the equations is based on the Verlet method[15], within which positions ($\boldsymbol{r}$), velocities ($\boldsymbol{v}$) and accelerations ($\boldsymbol{a}$) are expressed as Taylor expansions:

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \delta t \boldsymbol{v}(t) + \frac{1}{2}\delta t^2 \boldsymbol{a}(t) + ... \qquad [3.25]$$

$$\boldsymbol{v}(t+\delta t) = \boldsymbol{v}(t) + \delta t \boldsymbol{a}(t) + \frac{1}{2}\delta t^2 \boldsymbol{b}(t) + ... \qquad [3.26]$$

$$\boldsymbol{a}(t+\delta t) = \boldsymbol{a}(t) + \delta t \boldsymbol{b}(t) + \frac{1}{2}\delta t^2 \boldsymbol{c}(t) + ... \qquad [3.27]$$

The Verlet algorithm uses the positions and accelerations at time t and the positions from the previous step, to calculate the new positions at $(t+\delta t)$. This time step, δt, must be small enough to ensure that the system does not change significantly between t and $(t+\delta t)$ (usually δt ranges from 0.1fs to 10fs), in which case the solutions obtained are very similar to those that would be obtained solving the Newton's equations analytically. The trick of the Verlet algorithm is to add the positions at time $(t+\delta t)$ and $(t-\delta t)$, with the second order of approximation:

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \delta t \boldsymbol{v}(t) + \frac{1}{2}\delta t^2 \boldsymbol{a}(t), \qquad [3.28]$$

$$\boldsymbol{r}(t-\delta t) = \boldsymbol{r}(t) - \delta t \boldsymbol{v}(t) + \frac{1}{2}\delta t^2 \boldsymbol{a}(t) \quad . \qquad [3.29]$$

The sum of the equations [3.28] and [3.29] gives the final result:

$$\boldsymbol{r}(t+\delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\delta t) + \delta t^2 \boldsymbol{a}(t) \quad . \qquad [3.30]$$

Velocities are not needed for solving the equations, but they are needed for calculating thermodynamic properties. Velocities may be calculated dividing the difference in positions at different times:

$$v(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t}. \qquad [3.31]$$

This method is however not very precise. As can be seen from equation [3.30] r $(t+\delta t)$ is obtained by adding a small term (δt^2) to a difference of two large terms (O (δt^0)), which leads to loss of accuracy in the calculation. The Verlet leap-frog algorithm[9] does not have this problem; it is based on these equations:

$$r(t+\delta t) = r(\mathrm{t}) + \delta t v(\mathrm{t} + \frac{1}{2}\delta \mathrm{t}) \quad , \qquad [3.32]$$

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + \delta t a(t) \; + \mathrm{O}(\delta t^3), \qquad [3.33]$$

Velocities at time t may be calculated by:

$$v(t) = \frac{1}{2}\left[v(t+\frac{1}{2}\delta t) + v(t-\frac{1}{2}\delta t) \right]. \qquad [3.34]$$

Initially, the velocities leap over the positions to give their values at $(t+\frac{1}{2}\,\delta t)$. These velocities are more accurate than those obtained in the Verlet algorithm, having an error of the order of δt^3. Then the positions leap over the velocities to give their new values at $(t+\delta t)$, and so on. Another benefit in the accuracy of the Verlet leap-frog algorithm with respect to the simple Verlet consists of the fact that in the former algorithm the trajectory of the system at time $(t+\delta t)$ is obtained by adding a term of order δt, which is numerically more accurate than the trajectory at $(t+\delta t)$ in the simple Verlet algorithm, which is obtained by adding a smaller term of the order of δt^2 to a difference of large terms of the order of δt^0 (equation [3.30]).

The initial step in a molecular dynamics calculation is to assign the initial atom velocities distribution ($v(t-\frac{1}{2}\delta t)$ in equation [3.33]). Usually the velocities are randomly chosen from a Gaussian distribution and the Maxwell-Boltzmann distribution is rapidly established by molecular collisions within typically a few hundreds time steps. Therefore an atom i of mass m_i has a probability $p(v_{ix})$ of having a velocity v_{ix} in the x-direction at a temperature T, given by the Maxwell-Boltzmann distribution[16]:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \cdot e^{-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}} . \qquad [3.35]$$

The temperature is evaluated using the equipartition theorem of statistical mechanics:

$$K = \sum_{i}^{N} \frac{1}{2} m_i v_i^2 = \frac{3}{2} k_B T N \quad , \qquad [3.36]$$

where K is the kinetic energy and N the total number of particles in the system. Due to the random choice of initial velocities, it is necessary to run the simulation for a period long enough until the properties of the system do not change with time. When that is achieved the system is said to be equilibrated and time required for the system to equilibrate is the equilibration period. Results of the calculation are only analyzed after this period, in the so-called production period. Hence the velocities at time (t+1/2δt) and the forces at time t are calculated so that in the next step the trajectory of the system at time (t+δt) can be evaluated together with the velocities at time t. The procedure repeats iteratively.

3.5.2 Periodic Boundary Conditions (PBC)

For a relatively small system of molecules (~1000), arranged in a cubic cell, a large number of molecules lie close to the surfaces of the box. Molecules on the surface will experience different forces from those in the bulk, which is a problem when simulations of properties in the bulk are attempted. The problem of the surface effects can be overcome in MD simulations by implementing Periodic Boundary Conditions (PBC)[17], which allows us to model a macroscopic system in terms of a finite simulation system of N particles[9]. The box (which can be cubic or with other more complex geometries) is replicated throughout space to form an infinite lattice. In the course of the simulation, as a molecule moves in the original central box (in yellow, in figure 3.9), its "image" moves as well in the other boxes. At the same time, the boxes do not have any wall, so that, if a molecule in the central box moves away from it (molecule in blue) through a particular wall, its image simultaneously enters the central box from the opposite wall, as illustrated in figure 3.9.

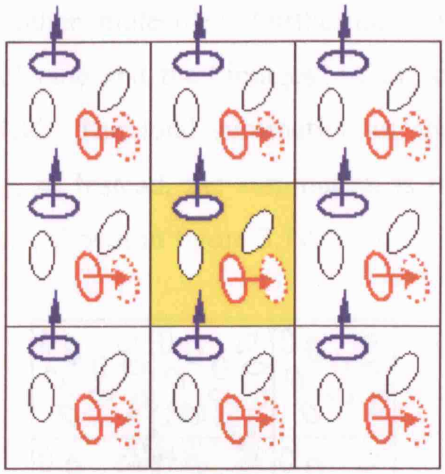


Fig. 3.9: Illustration of the application of *periodic boundary conditions* to a central box (in yellow) of a MD simulated periodic macroscopic system. The arrows indicate the movement of the molecules. The red dotted line molecules represent the new positions of the translated red molecules.

If the particle in red in the central box is located at position vector $\boldsymbol{r}$ in a box, the identical particles are assumed to be located at the positions:

$$\boldsymbol{r'} = \boldsymbol{r} + (i\boldsymbol{A} + j\boldsymbol{B} + k\boldsymbol{C}) \quad , \qquad [3.37]$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are the three vectors defining the box of the simulation. The integers i, j and k vary from $+\infty$ and $-\infty$. It is important to understand the influence of the PBC on the physical properties of the simulated system. Some particles, which can be considered far away from other particles within a box, can be very close to periodic repeats of these particles in adjacent boxes, so that their interactions can inhibit the occurence of long wave-length fluctuations and prevent the simulation of first order phase transitions, such as gas-liquid, in which the critical fluctuations become macroscopic near the phase coexistence. The influence of PBC on the simulation is also related to the potential used. In the case of a long-range potential, of the form $V(r) \sim r^{-\nu}$, with ν being less then the dimensionality of the system, there will be a substantial interaction between a particle and its own images in neighbouring boxes and consequently the symmetry of the cell is imposed in the system, with the consequent large influence of PBC on the simulation. For simulated systems far from the the phase coexistence, the PBC exert no strong influence on the calculated properties[9], for box lengths $L < 10\sigma$, indicating with σ the atomic diameter . The system size strongly influences the pressure fluctuations[18].

In calculating the forces acting on a certain molecule (for instance the green molecule in figure 3.10) in a box containing N molecules, we must include interactions between this

molecule and all the N-1 other molecules. Furthermore we should also include the interaction between this molecule and the "images" of the other molecules in the other boxes, generated by the PBC. The total summation of the forces in this latter case contains an infinite number, so instead, the summation is restricted by considering the "*minimum image convention*", shown in figure 3.10.

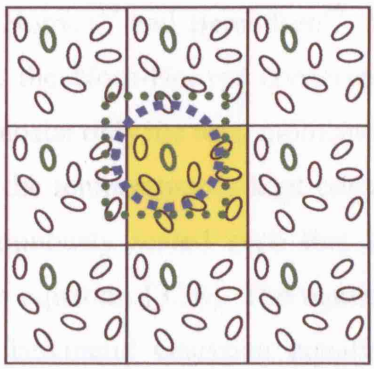


Fig. 3.10: "*Minimum image convention*" region (green dotted line) and cut-off sphere (blue dotted line), in calculating the potential energy and the forces of a specific molecule (green molecule) with all the other molecules in the system.

A region is defined about each molecule, which has the same shape and size as the MD box (figure 3.10: green dotted line) and the potential energy is calculated from the interactions of the central molecule and the N-1 neighbours in this box. By using this convention, the evaluation of the potential energy of the system within the pair-wise approximation involves $\frac{1}{2}N(N-1)$ terms. A further approximation consists of considering the largest contribution to the potential energy (and forces), which comes from neighbours close to the chosen molecule within a spherical cut-off of radius r_c. For distances larger than r_c the pair potential and the forces are set to zero. The spherical region is co-centric with the "minimum image" region (figure 3.10: light blue dotted line) and for consistency with the latter the spherical cut-off radius must be no greater than half of the length of the box dimension:

$$r_c \leq \frac{1}{2}L \quad , \qquad [3.38]$$

where L is the dimension of a cubic box. Of course, in the case of non cubic boxes, this condition must be applied to the shortest of the three vectors defining the MD simulation box.

3.5.3 Thermostats & Barostats

There are several statistical ensembles, which we can use in MD simulation, which are discussed in terms of which quantities are conserved: these include the NVE, NVT, NPT, ensembles. In our studies we mainly use the NVE and anisotropic constant pressure NPT ensembles (NST in DL_POLY_2 notation). For the latter we attempted to use two different thermostats: Nose-Hoover[19] and Berendsen[20]. The essential difference between the two thermostats is that the Nose-Hoover conserves momentum and total energy, while in the Berendsen thermostat only the total momentum is conserved.

In the *isokinetic* thermostat the temperature is kept constant under the whole simulation and the velocities are continuously scaled such that the total kinetic energy is kept constant, in conformity with equation [3.36]. The equations describing the dynamics of the system are known as isokinetic Gaussian equations of motion, and they were proposed by Hoover et al[19,21]. In the Nose-Hoover NVT thermostat, the temperature is not kept constant, but is allowed to oscillate. In this case the modified Hamiltonian equations of motion are:

$$\frac{d\boldsymbol{r}_i(t)}{dt} = \boldsymbol{v}_i(t), \qquad [3.39]$$

$$\frac{d\boldsymbol{v}_i(t)}{dt} = \frac{\boldsymbol{f}_i(t)}{m_i} - \alpha(t)\boldsymbol{v}_i(t) \quad , \qquad [3.40]$$

where the frictional parameter $\alpha(t)$ is controlled by the following first order differential equation:

$$\frac{d\alpha(t)}{dt} = \frac{1}{\tau_T^2}\left(\frac{T}{T_0} - 1\right) \quad , \qquad [3.41]$$

where τ_T is the relaxation time of the thermostat (normally in the range between 0.5ps and 2ps), which has to be assigned in the input of any calculation in order to minimize the fluctuations of the temperature in the system. T_0 is the external temperature of the system, fixed in the input, while T is the local oscillating temperature. The consequence of the time variation of the frictional factor is that there is an influence over the system in order to change its temperature, making it fluctuate during the whole simulation.

In the Berendsen[20] algorithm the temperature is pushed towards the desired one (T_0), by scaling the velocities at each step, with the scaling factor α. Hence the approach is different from the Nose–Hoover one, in which the Hamiltonian equations of the motion are modified. The scaling is given by the equations:

$$\alpha \leftarrow \left[1+\frac{\Delta t}{\tau_T}\left(\frac{T}{T_0}-1\right)^{1/2}\right], \qquad [3.42]$$

$$v(t+\frac{1}{2}\Delta t) \leftarrow \left[v(t-\frac{1}{2}\Delta t)+\Delta t\frac{f(t)}{m}\right]\cdot\alpha\ , \qquad [3.43]$$

$$v(t) \leftarrow \left[v(t-\frac{1}{2}\Delta t)+v(t-\frac{1}{2}\Delta t)\right]\cdot\alpha\ , \qquad [3.44]$$

$$r(t+\Delta t) \leftarrow r(t)+\Delta t v(t+\frac{1}{2}\Delta t)\,. \qquad [3.45]$$

The fluctuation of the pressure has to be also considered in the case of working an NPT ensemble. The size and/or the shape (in NPT ensemble) of the system can be dynamically adjusted, by coupling it with a barostat in order to obtain a desired average pressure (the external pressure P_0) and/or an isotropic pressure/stress tensor ($\boldsymbol{\sigma}$). Note that Berendsen thermostat does not sample the canonical ensemble. With the Berendsen barostat the system the system is made to obey to the equation of motion:

$$\frac{dP}{dt}=\frac{1}{\tau_P}(P_0-P)\ . \qquad [3.46]$$

P is the instant, fluctuating, pressure in the MD simulation and τ_P is the relaxation time for the barostat, which has to be assigned as input parameter for the MD simulation. In the isotropic implementation, at each step, the MD cell volume is scaled by a factor η and coordinates and cell vectors are scaled by a factor $\eta^{1/3}$. The scaling factor, also denoted barostat friction coefficient, is defined as[10]:

$$\eta = 1-\frac{\beta\Delta t}{\tau_P}(P_0-P)\,. \qquad [3.47]$$

The constant β is the isothermal compressibility of the system (DL_POLY[10] considers as convention the value of the compressibility β of the liquid water). Hence, after having determined the new coordinates in equation [3.45], the scaling of the atomic positions and box size is performed:

$$r(t+\Delta t) \leftarrow \eta^{1/3} r(t+\Delta t)\,, \qquad [3.48]$$

$$l(t+\Delta t) \leftarrow \eta^{1/3} l(t+\Delta t)\,, \qquad [3.49]$$

$$V(t+\Delta t) \leftarrow \eta V(t+\Delta t)\,. \qquad [3.50]$$

The extension of equation [3.47] to the case of anisotropic constant pressure ensemble (NST), in which the cell can change both size and shape, is straightforward:

$$\boldsymbol{\eta} = \mathbf{1} - \frac{\beta \Delta t}{\tau_P}(P_0 \mathbf{1} - \boldsymbol{\sigma}), \qquad [3.51]$$

where the notation in bold indicates a tensor. The new cell vectors are given by the tensorial scaling factor $\boldsymbol{\eta}$:

$$\boldsymbol{H}(t + \Delta t) \leftarrow \boldsymbol{H}(t) \cdot \boldsymbol{\eta}, \qquad [3.52]$$

where $\boldsymbol{H}$ is the matrix formed by the three cell axes, $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.

$$\boldsymbol{H} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \quad . \qquad [3.53]$$

3.5.4 Statistical Quantities: radial distribution function & mean square displacement

Two important tools used in our work, which can be calculated from an MD simulation are the Radial Distribution Function (RDF) *g(r)* and the mean square displacement (MSD). The RDF (sometimes denoted "pair distribution function"), which can be measured experimentally via X-ray and neutron diffraction, is a measure of the local environment of the atoms in the system, and it is the probability of finding two species of atoms separated by a distance $\boldsymbol{r}$. More precisely, the pair radial distribution function, $g(r_1, r_2)$ or $g(r_{12})$ or $g(r)$, gives the probability of finding a pair of atoms a distance $\boldsymbol{r}$ far apart, with respect to the probability expected for a completely random distribution at the same density. In order to define mathematically this function we integrate the configurational distribution function over all the atoms in the system apart from the two we are taking in consideration. In the canonical ensemble (NVT) the expression for the pair radial distribution function is:

$$g(r_1, r_2) = \frac{N(N-1)}{\rho^2 Z_{NVT}} \int e^{-\beta V(r_1, r_2, r_3, r_4 \ldots . r_N)} d\boldsymbol{r}_3 d\boldsymbol{r}_4 \ldots . . d\boldsymbol{r}_N \quad . \qquad [3.54]$$

In equation [3.54], N is the total number of atomic species in the system, ρ is the density of the system, the subscripts *1, 2, ...N* refer to the different atomic species in the system, and Z_{NVT} is the configurational integral of the system:

$$Z_{NVT} = \int e^{-\frac{V(r)}{k_B T}} d\boldsymbol{r} \quad . \qquad [3.55]$$

Typical radial distribution functions for a liquid and a solid are shown in figure 3.11.

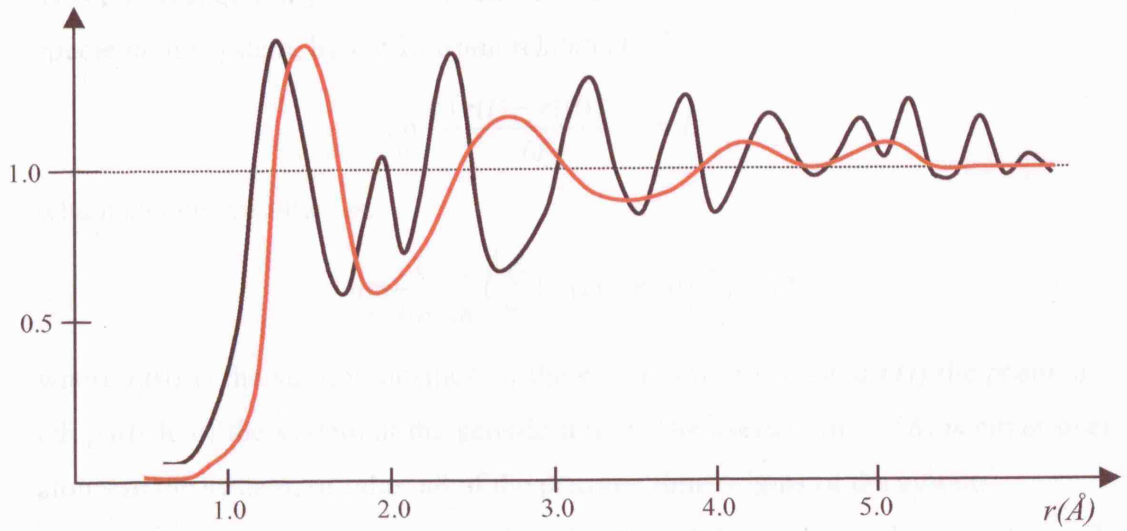


Fig. 3.11: Examples of radial distribution functions (RDF) for a solid crystalline phase (black curve) and a liquid disordered phase (red curve).

For a crystalline phase, characterized by the presence of both long and short – range order, the RDF is very structured, including for large distances between atoms (figure 3.11: black curve). By contrast, in the liquid state, only short range order is reflected in the presence of peaks at small values of distance, whilst at longer distance $g(r)$ smoothes to the value of 1 (figure 3.11: red curve).

In the case of molecules, the pair radial distribution function $g(\boldsymbol{r}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j)$ will also depend on the orientation of the two molecules, $\boldsymbol{\Omega}_i$, $\boldsymbol{\Omega}_j$. A site-site description is appropriate in describing this RDF of the system. Pair radial distribution functions $g_{ab}(\boldsymbol{r}_{ab})$ are defined for each pair of sites on different molecules (g_{ab} for CC, CH, NH, NN and so on), using the same definition illustrated for the atomic case. They have the advantage of being directly related to the structure factor of the solid, hence to the experimentally observable properties. The pair–pair RDF can be easily computed from the trajectory file of the systems, which contain the positions of the individual atoms in time. For a specific pair - pair RDF, the final function is evaluated as a time average over all the different configurations of the system.

The mean square displacement $<\boldsymbol{r}^2>$ (*MSD*) of the system is the average of the square of the displacements of the molecules, during the time of simulation. It can be easily calculated from the atomic positions in the trajectory file of the simulation, using the formula[8]:

$$MSD = <\boldsymbol{r}^2> = <[\boldsymbol{r}(t) - \boldsymbol{r}(0)]^2> . \qquad [3.56]$$

This physical quantity is directly related to the diffusion coefficient D of a certain atomic specie in the system, by the Einstein relationship[8]:

$$\lim_{t\to\infty}\frac{<[r(t)-r(0)]^2>}{6t}=D\ , \qquad [3.57]$$

which can be rewritten as:

$$\lim_{t\to\infty}\frac{1}{6N}\frac{d}{dt}\left\langle\sum_{i=1}^{N}[r_i(t)-r_i(0)]^2\right\rangle=D\ , \qquad [3.58]$$

where $r_i(0)$ is the starting position of the system (at time 0) and $r_i(t)$ the position of the i-eth particle of the system at the generic time t. The average in [3.58] is either over all the atoms in the system, or other all of the possible time origins of the system[8].

The typical trends of the *MSD* for a liquid and a solid system are shown in figure 3.12.

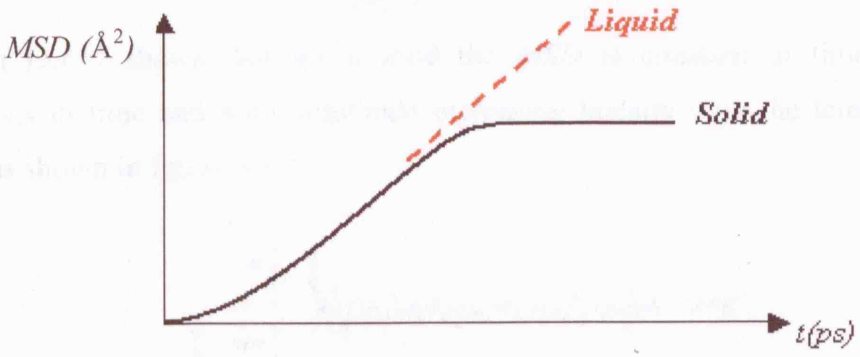


Fig. 3.12: Schematic trend of the mean square displacement for a liquid and a solid phase.

At very short times, the plot is not linear, because the path followed by a molecule will be an approximate straight line until it collides with its neighbour. Only when it starts the collision process, will its path begin to resemble a random walk. Until it makes that first collision, we may say it moves with an approximately constant velocity, i.e. the distance, which it travels, is proportional to time, and its *MSD* is therefore proportional to the time squared. Thus at very short time, the *MSD* trend resembles a parabola.

For a solid crystalline phase, characterised by phonon vibrations treated within the harmonic approximation, the *MSD* amplitude in the limit of time $t\to\infty$, is given by the general expression[22,23]:

$$<r^2>=\frac{h}{8\pi^2\mu\nu}coth\left(\frac{h\nu}{k_BT}\right)\ , \qquad [3.59]$$

where ν is the characteristic frequency of the phonon vibration, μ is the reduced mass of the system and *h* is the Planck constant. This equation shows an oscillating behaviour of the *MSD* around an average constant value in time. For very big values of the term $\frac{h\nu}{k_B T} >> 1$ (i.e. at very low temperatures) equation [3.59] reduces to that for zero-point motion:

$$< r^2 >= \frac{h}{8\pi^2 \mu \nu} \qquad [3.60]$$

For very small values of the term $\frac{h\nu}{k_B T} << 1$, the expression in [3.59] reduces to the classical expression:

$$< r^2 > \sim \frac{k_B T}{\mu (2\pi\nu)^2} \qquad [3.61]$$

Equation [3.61] shows that for a solid the *MSD* is constant in time, subjected to oscillations in time and with amplitude increasing linearly with the temperature of the system as shown in figure 3.13.

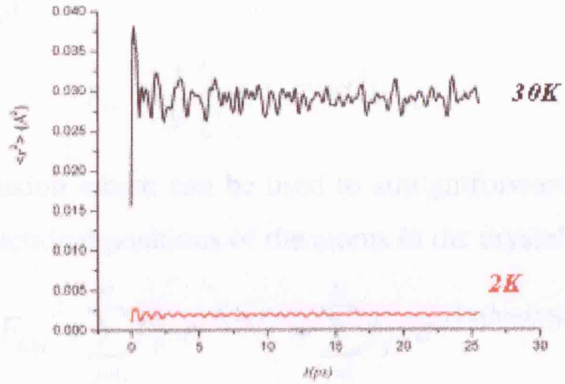


Fig. 3.13: Mean square displacement (<r^2>) of the MD simulated solid cyclopentane at very low temperatures, of 2K (red) and 30K (black).

3.6 Calculation of X-Ray Powder Pattern from MD simulated system

Electrons and other charged particles in a crystal can scatter an X-Ray beam. The interference of their scattering, together with the periodicity of the crystal results in the diffraction of the radiation, which produces a diffraction pattern, composed of peaks of different intensities at different angles. The diffraction pattern contains all the information concerning the symmetry of the crystal, from which we can extract atomic coordinates and structural parameter for the unit cell.

Our aim is to calculate the integrated intensities of the diffraction pattern for a molecular dynamics (MD) simulated crystal, and to study the way the pattern evolves with temperature. The simplest expression for the integrated intensity of a peak at an angle, θ, is[24]:

$$I_{hkl}(\theta) = |F_{hkl}|^2 \cdot C(\theta), \qquad [3.62]$$

where F_{hkl} is the structure factor and $C(\theta)$ is a function, which contains factors dependent on the angle of the incident beam, which will be treated more in detail below.

The real source of radiation scattering in the crystal is the electron density of the atoms in the structure[25,26], which in a crystal can be expressed with a Fourier expansion as in equation [3.63]:

$$\rho(\boldsymbol{r}) = \sum_{\boldsymbol{hkl}=-\infty}^{+\infty} F_{hkl} e^{i\boldsymbol{G}_{hkl}\cdot\boldsymbol{r}} \quad , \qquad [3.63]$$

with $\boldsymbol{G}_{hkl}$ being the reciprocal lattice vector and *hkl* representing the Miller indexes, which specify the position of a point in the reciprocal lattice. Hence, by doing a Fourier trnasformation of the expression [3.63], it becomes possible to calculate the structure factor for a crystal in a reciprocal lattice, from the electron density of the atoms and their positions in the crystal:

$$F_{hkl} = \frac{1}{V}\int_0^r \rho(\boldsymbol{r}) e^{-i\boldsymbol{G}_{hkl}\cdot\boldsymbol{r}} d\boldsymbol{r} \quad , \qquad [3.64]$$

An alternative expression which can be used to straightforwardly calculate the structure factor from all the fractional positions of the atoms in the crystal is:

$$F_{hkl} = \sum_{j=1}^{N} f_j \cdot e^{-i\boldsymbol{G}_{hkl}\cdot\boldsymbol{r}_j} = \sum_{j=1}^{N} f_j \cdot e^{-i2\pi(hx+ky+lz)} \quad , \qquad [3.65]$$

with *x, y, z* being the fractional atomic coordinates and *N* the total number of the atoms in the crystal, and f_j being the *atomic scattering factor* or *form factor*[27], which represents the ratio of scattering from the atom to that of a single electron. The form factor is given by:

$$f_j = \int_V \rho(r - r_j) e^{-iG\cdot(r-r_j)} dV \quad , \qquad [3.66]$$

where V is the volume of the unit cell and the integral is over all the volume of the system. The form factor is proportional to the atomic number (Z) of certain atomic species and depends also on the angle (θ) of the diffracted beam with wave-length λ, according to the following expression:

$$f_j = Z - 41.78214 \cdot \frac{sin^2(\theta)}{\lambda^2} \cdot \sum_{i=1}^{4} a_i e^{-b_i \frac{sin^2\theta}{\lambda^2}} \quad . \qquad [3.67]$$

In this formula a and b are coefficients, which depend on the kind of atoms. The angular dependence of f_j is illustrated in figure 3.14. As we can infere from equation [3.67], at very low angles f_j is only proportional to the atomic number Z.

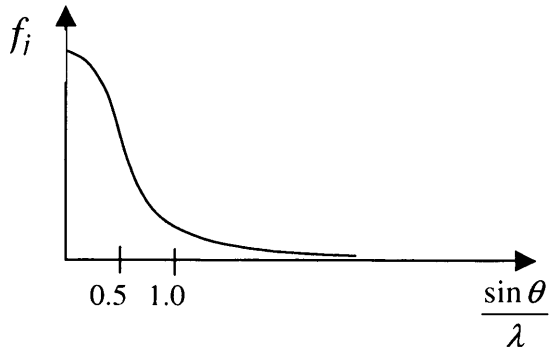


Fig. 3.14: Trend of the form factor as function of angle of the diffracted beam[5].

Using equation [3.65] we can calculate theoretically the structure factor for a certain number of discrete values of the three Miller indexes, corresponding to the values of the angle 2θ. Equation [3.65] or [3.64] has been the starting point of the calculation of powder patterns from MC and MD simulated systems in other previous works[28,29]. However this is not sufficient for a good calculation of the integrated intensity of the diffraction peaks.

To calculate integrated intensities we have to take in account two essential factors[30] affecting the amplitude of a diffraction peak:

a) The Lorentz-polarization factor (*Lp*)

b) The Temperature factor

The *Lp* factor is due to the fact that the efficiency of the electrons in scattering the radiation changes with the angle of the incident beam. It is strongest for forward and back scattering, while it is zero along a direction parallel to the electric field of the radiation. The Thomson equation giving an estimation of the polarization factor is:

$$I_p \propto \frac{1+cos^2\theta}{2} \quad . \qquad [3.68]$$

At the same time, the integrated intensities also depend on the height and the width of the curve of the peaks in the pattern, which in turn are related to the Bragg angle and to the distribution of the grains in the powder. The Lorentz factor is given by the expression:

$$I_L = \frac{1}{2sin^2\theta \cdot cos\theta} \quad . \qquad [3.69]$$

Combining these two contributions, we obtain the *Lp* multiplicative factor for the integrated intensity of a diffraction peak;

$$Lp = \frac{1+cos^2\theta}{2(sin\theta \cdot sin2\theta)} \quad . \qquad [3.70]$$

The temperature factor results from the oscillatory vibration of the atoms around their centre of mass. The vibrational motion leads to a further angular dependence contribution to the diffracted peak intensity, which has the following expression:

$$I_T \propto e^{-2M} \quad , \qquad [3.71]$$

where

$$M = B_j \cdot \left(\frac{sin\theta}{\lambda}\right)^2 \quad . \qquad [3.72]$$

The coefficient B_j is called *"isotropic temperature factor"* and is proportional to the atom's square displacement about the atomic equilibrium position. In figure 3.15a and 3.15b we show these factors as functions of the diffraction angle (θ).

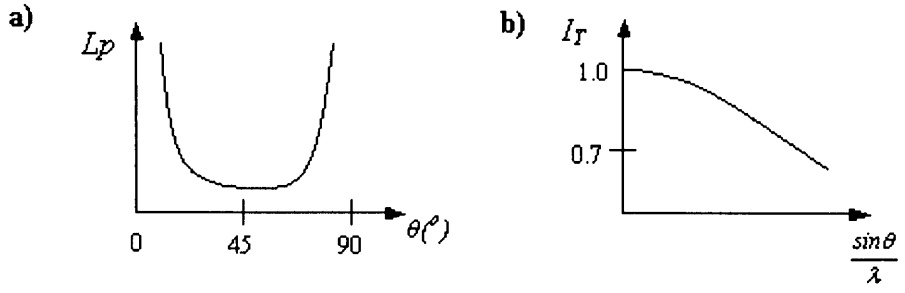


Fig. 3.15: a) Lorentz-polarization factor as function of θ. **b)** Temperature factor as of θ.

Taking in account these two main angular contributions we can write the integrated intensity with using the following expression:

$$I_{hkl}(2\theta) = |F_{hkl}|^2 \cdot Lp \cdot I_T \quad . \qquad [3.73]$$

The outcome of the calculation by using the formula [3.73] is a spectrum with sharp lines corresponding to the peaks diffracted at specific angles by the crystal structure, as shown in figure 3.16a.

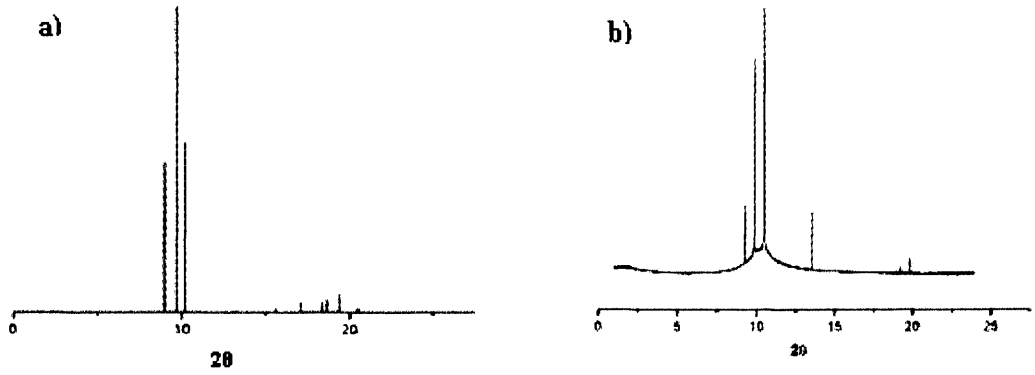


Fig. 3.16: a) Calculated structure factor for cyclopentane at 140K. **b)** Experimental X-Ray spectrum for cyclopentane at 140K. The intensities in the patterns are not normalised.

An experimental spectrum is characterised by having slightly broader diffraction peaks (figure 3.16b), due to several microstructural parameters for example background and instrumental resolution, specimen defects and preferred orientations. In this latter case, the broadening is given by the presence of a shape anisotropy of the particles of the sample, which are not truly randomly distributed in the powder. To simulate properly the peak profile in a powder pattern we have to consider the convolution of different functions to model this broadening effect of the peaks. The standard way to proceed is to consider pseudo-Voigt functions[31,32] centred on each peak of the calculated structure factor. These functions are linear combinations of Gaussian functions $G(x)$ and Cauchy functions $C(x)$, with the form:

$$C(x) = \frac{1}{1+x^2} \quad , \qquad [3.73]$$

$$G(x) = e^{-ln2 \cdot x^2} \quad , \qquad [3.74]$$

with

$$x = \frac{(2\theta - 2\theta_o)}{w} \quad , \qquad [3.75]$$

where $2\theta_o$ is the angular value at which every peak is centred (the position of the peak maximum) and w is the half of the full width at half-maximum of the peak (FWHM). The latter parameter determines the broadening of the peak and ususally is set up around the value of 0.02-0.03, in order to model appropriately the experimental broadening.

Hence for each value of positions of the maxima ($2\theta_i$), at which we have calculated the structure factor we construct a pseudo-Voigt function in order to obtain the final theoretical X-Ray powder pattern having this analytical expression:

$$I_{Calc}(\theta) = \sum_{i=1}^{M} \left[\eta \cdot \frac{1}{1+x_i^{\,2}} + (1-\eta) \cdot e^{-ln2 \cdot x_i^{\,2}} \right] \cdot I(\theta_i) \quad . \qquad [3.76]$$

We indicate with M the total number of angles, for which we calculated the theoretical structure factor. The parameter η is known as the Cauchy content and usually is set up to the standard value 0.5, which consider equal contribution of Cauchy functions and Gaussian functions in modelling the peak shape. The method is shematically visualised in figure 3.17.

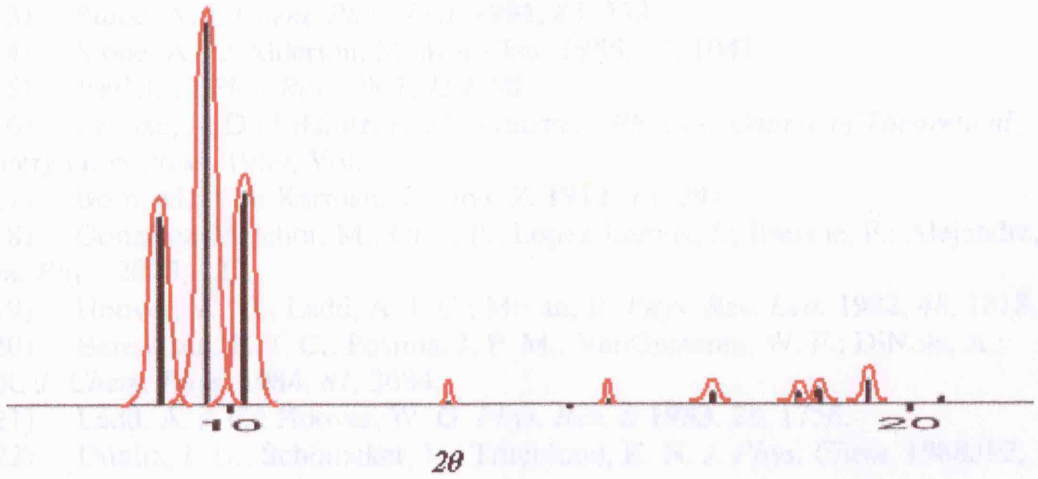


Fig. 3.17: Schematic representation of the pseudo-Voigt function convolution of the calculated structure factor in the simulation of a X-Ray pattern.

3.7 References

(1) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications: New York, 1996.

(2) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(3) Angyan, J. G.; Chipot, C. *Int. J. Quantum Chem.* **1994**, *52*, 17.

(4) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(5) Gdanitz, R. J. *Chem. Phys. Lett.* **1992**, *190*, 391.

(6) Holden, J. R.; Du, Z. Y.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.

(7) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628.

(8) Haile, J. M. *Molecular Dynamics Simulation:Elementary Methods*; Wiley & Sons: United States, 1992.

(9) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquid*; Oxford Science: Oxford, 1987.

(10) Smith, W.; Forester, T. R. *J.Mol.Graphics* **1996**, *14*, 136.

(11) Gray, A. E.; Day, G. M.; Leslie, M.; Price, S. L. *Mol. Phys.* **2004**, *102*, 1067.

(12) Price, S. L.; Hamad, S.; Torrisi, A.; Karamertzanis, P. G.; Leslie, M.; Catlow, C. R. A. *Mol. Simul.* **2006**, *32*, 985.

(13) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(14) Stone, A. J.; Alderton, M. *Mol.Phys.* **1985**, *56*, 1047.

(15) Verlet, L. *Phys Rev.* **1967**, *159*, 98.

(16) Landau, L. D.; Lifshitz, E. M. *Statistical Physics: Course of Theoretical Physics*; Pergamon Press, 1969; Vol. 5.

(17) Born, M.; Von Karman, T. *Phys. Z.* **1912**, *13*, 297.

(18) Gonzalez-Melchor, M.; Orea, P.; Lopez-Lemus, J.; Bresme, F.; Alejandre, J. *J. Chem. Phys.* **2005**, *122*.

(19) Hoover, W. G.; Ladd, A. J. C.; Moran, B. *Phys. Rev. Lett.* **1982**, *48*, 1818.

(20) Berendsen, H. J. C.; Postma, J. P. M.; VanGusteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(21) Ladd, A. J. C.; Hoover, W. G. *Phys. Rev. B* **1983**, *28*, 1756.

(22) Dunitz, J. D.; Schomaker, V.; Trueblood, K. N. *J. Phys. Chem.* **1988**, *92*, 856.

(23) Sales, B. C.; Chakoumakos, B. C.; Mandrus, D.; Sharp, J. W. *J. Solid State Chem.* **1999**, *146*, 528.

(24) Rohrer, G. S. *Structure and Bonding in Crystalline Materials:239*Cambridge, 2001.

(25) Rohrer, G. S. *Structure and Bonding in Crystalline Materials:206-208*Cambridge, 2001.

(26) Rohrer, G. S. *Structure and Bonding in Crystalline Materials:220;235*, 2001.

(27) Rohrer, G. S. *Structure and Bonding in Crystalline Materials:220-235 - Scattering atomic factor*Cambridge, 2001.

(28) McGreevy, R. L. *J. Phys.-Condes. Matter* **2001**, *13*, R877.

(29) Kihara, K.; Matsui, N. *Phys. Chem. Miner.* **2003**, *30*, 345.

(30) Rohrer, G. S. *Structure and Bonding in Crystalline Material: 236*Cambridge, 2001.

(31) Sanchez-Bajo, F.; Cumbrera, F. L. *J. Appl. Crystallogr.* **1997**, *30*, 427.

(32) Langford, J. I. *J. Appl. Crystallogr.* **1978**, *11*, 10.

Chapter 4

IMIDAZOLE CRYSTAL STRUCTURE PREDICTION

4.1 Introduction

This chapter describes a work on crystal structure prediction (CSP) for polymorphs of organic molecular crystals, which is an important area of computer modelling.

Polymorphism is one of the most difficult and significant problems involving molecular crystals[1]. In particular, from the computational point of view, the prediction of the different polymorphs and their related energies is a difficult and delicate issue[2,3]. Hence it is necessary that the methods used are as accurate as possible.

The molecule studied in this work is imidazole. It is a relatively simple organic molecule, which is used in a wide variety of chemical fields, as in the production of pharmaceuticals, pesticides, dye intermediates, photographic chemicals and corrosion inhibitors. In figure 4.1 is shown the structure formula of the molecule.

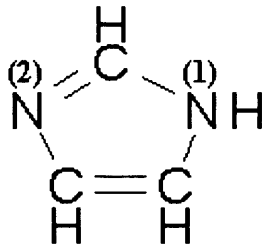


Fig. 4.1: Imidazole Structural Formula. The numbers (1) and (2) indicate the different types of nitrogen atoms in the molecule.

The molecular shape is a pentagonal ring with the hydrogen bonded to nitrogen more polarized compared to the other hydrogens. The molecule is close to being planar, due to the presence of two double bonds in the ring structure and a consequent rigidity.

One crystal structure of this molecule is known and experimentally investigated through neutron diffraction and reported in the Cambridge Structure Database[4] (CDS).

The purpose of this work is first to discuss the results obtained from the lattice energy minimisation of this crystal structure and to compare them with the results of a previous investigation[5]. Our work, however, goes beyond the earlier studies, by analysing the

differences between the minimised lattice-energies for the structure with the molecule in the experimental conformation and the one with the molecule in gas phase, *ab-initio* calculated, conformation molecular structure, which is needed for the crystal structure prediction as starting molecule used in the theoretical method to build all the possible polymorphs (chapter 3 section *3.3.2*).

Subsequently we performed a CSP of imidazole to examine whether the experimental structure corresponds to the predicted with lowest lattice energy and to investigate the possibility of polymorphs for crystals of this molecule.

4.2 Methods

The data on the CSD for the imidazole derive from a neutron diffraction experiment[6] with a low R-factor of 0.04, determined at the temperature of 103K.

As previously stated, it makes sense to treat the molecule as a rigid unit building the crystal lattice.

4.2.1 Molecular Electronic Structure & Minimisation Energy techniques

In the first minimisation the molecular geometry is that from the experimental data reported in CSD. In this simulation we did not adjust the bond length of the C-H and N-H groups to the typical values of 1.08 Å and 1.01 Å respectively, because of the high accuracy of the neutron diffraction technique in determining the hydrogen nuclear positions.

The second minimisation procedure starts with an optimization of the molecule geometry, using an *ab initio* quantum approach, which calculates the minimum energy structure, treating the molecule as an isolated gas phase species. The used method is an MP2 corrected Hartree Fock[7] implemented in GAUSSIAN98 software package[8] with an MP2/6-31G** wave function[9] (section 3.2).

The calculations use interatomic potentials, which were previously discussed in section 1.5. The values of the potential parameters A_{tt}, B_{tt}, C_{tt}, in the 6-exp potential form are are those of the FIT set potential[10] (table 3.2 in appendix 3.4, chapter 3). This set of parameters is the same used in a previous study of the imidazole crystal structure, which has shown the good quality of the potential for such compound[5].

The electrostatic contribution to the intermolecular interactions was treated through a multipole expansion of the molecular density charge, using a distributed multipole analysis (DMA[11]: section 1.4.2). The atomic point charges in the molecule were derived from the electrostatic potential (ESP) using the CHelpG algorithm[12] (section 3.1), which estimates their values by fitting the atomic charges to the molecular electrostatic potential surrounding the molecule. However, the presence of the hydrogen bond (HB) between (N)H•••N(C) in the molecular crystal of the imidazole and the presence of some electronic features of the imidazole molecule (π - electron density) is not adequately described by a spherical atom model of the electrostatic potential. The multipole expansion of each single atomic charge provides charge, dipole, quadrupole, octupole, hexadecapole.

We calculated the lattice energy minimisation using either the DMA or using only the atomic point charges calculated with CHelpG technique. This was made either for the experimental molecular geometry or for the optimised one.

The electrostatic energy contribution involved in the intermolecular interactions was calculated through a multipole expansion up to the fifth order of the distance (R^{-5}).

The use of the FIT parameters for the intermolecular potential and the DMA for the multipole expansion was found in previous calculations to give the best results in the simulation of the crystal structure for the different kinds of molecules included imidazole[5,13] (APPENDIX D).

The program used for minimising the lattice energy is DMAREL[14], previously discussed in chapter 3, section 3.4. The charge-charge, charge-dipole, dipole-dipole electrostatic contributions to the lattice energy are evaluated by the Ewald summation technique[15], discussed in appendix A, with the relative values of cut-offs for the sum in the direct and reciprocal spaces given respectively by the equations [A.13] and [A.14], in APPENDIX A. All higher order terms in the multipole expansion are summed to a 15Å cut-off between the centres of mass of the molecule. For the repulsion-dispersion terms an atom-atom separation of 15 Å was used as cut-off .

The local molecular axis is defined so that the *x* axis passes through the two nitrogen atoms in the molecule, the *y* axis forming a plane with *x* axis defined by the atoms N1, N2, and C3 and the *z* axis perpendicular to this plane.

4.2.2 *Crystal Structure Prediction Procedure*

The second step of our work involved the prediction of all the possible structures of imidazole, starting from the molecule in the optimised geometry (or in the gas phase).

The crystal prediction was performed using MOLPAK[16], a program generating all the possible crystal structures for organic molecules, containing the C, H, O, F, N elements, in the more common space groups adopted in organic solid state (section 3.3).

For each packing type, MOLPAK performs a systematic search for dense packings, building a maximum number of 200 structures for each packing type.

After the construction of all the possible packing types, DMAREL[14] program automatically minimises the lattice energy of each packing type. The final structures are listed in a table, in which for each structure we report the packing type, the initial and final lattice energy, initial and final volume of the unit cell, the lattice parameters of the unit cell and the lattice parameters of the reduced cell.

Using appropriate programs (PostProp, PostSearch and PLATON[17]) the structures were then sorted. For a given structure, we remove all of the similar structures, which have differences in lattice parameters less than specific threshold values, which distinguish two different structures, within the accuracy of the program:

ΔEnergy = 0.1 kJmol^{-1}

ΔVol = 0.1 Å^3

Δ**a** (**b**,**c**) = 0.01Å

$\Delta\beta$ = 0.05 °.

For the final structures, the lattice energy *vs* volume/molecule are plotted in a graph (scatterplot), in order to give a picture of the lattice energy landscape of the system.

We also calculated the phonon frequencies for both of the imidazole experimental minimised crystal structures, one with the molecule in the experimental conformation and one with the molecule in the *ab-initio* optimised geometry; we also calculate the phonon frequencies for the predicted structures in the range of 5kJmol^{-1} above the global minimum (GM). Both the elastic constants and the phonon frequencies were calculated within the harmonic approximation (described in appendix 3.2, chapter 3) neglecting the temperature effects, which has been shown to predict accurate values of these quantities, when compared with experiment[1].

4.3 Results and Discussion

4.3.1 Molecular calculations

The experimental crystal lattice of imidazole has a monoclinic unit cell with space group $P2_1/c$. The *ab-initio* optimised geometry of the imidazole molecule is reported in figure 4.2.

Fig. 4.2: Imidazole "gas phase" optimised geometry.

Table 4.1 reports the results from the calculations of the molecular electronic structure for both the experimental and the optimised molecules.

	***E(SCF)* /Hartree**	***E(MP2)* /Hartree**	***Dipole moment* /Debye**	***C-N(1)-C angle* /o**	***C-N(2)-C angle* /o**	***H-C-N(1)-H torsion* /o**
Exp molec	-224.8199	-225.5505	3.92	107.1	105.1	2.24
Opt molec	-224.8213	-225.5525	3.86	107.6	104.9	0.14

Table 4.1: *Ab-initio* calculated electronic properties of the imidazole molecule, in the experimental (*Exp molec*) and in the *ab-initio* optimised (*Opt molec*) geometry.

The optimised and experimental molecular geometries differ only by 6 $kJmol^{-1}$, with the optimised molecular geometry being slightly different than the experimental, with the N-H intra-molecular bond moved into the same plane of the molecular pentagonal ring. It is worth underlining the good agreement between the values of the molecular dipole moment, with a small decreasing of the dipole moment of the optimised molecule with respect to the experimental one.

Table 4.2 reports the point charges obtained from this method for each atom and figure 4.3 shows the electron density distribution.

Atom type	*Charges /e*
N1	-0.23
N2	-0.51
C1	0.22
C2	0.18
C3	-0.24
H1	0.30
H2	0.07
H3	0.05
H4	0.16
Total Dipole Moment = 3.795 Debye	

Table 4.2: CHelpG calculated atomic point charges for imidazole *ab-initio* optimised molecule.

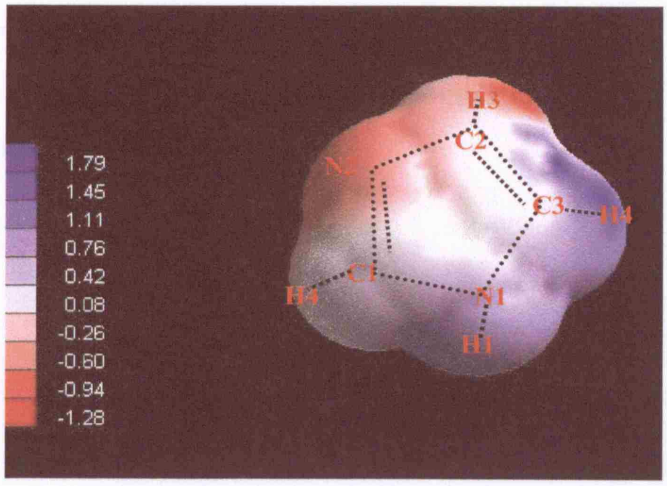


Fig. 4.3: CHelpG electron density distribution around the *ab-initio* optimised imidazole molecule. As indicated by the legend of the colours, the negative charge is localised in the areas in red around the molecule, while the positive charge in the areas in blue.

For the optimised molecule, the dipole moment, calculated by fitting the charges to the potential, is in good agreement with that directly obtained from the wave function and reported in table 4.1, which confirms the good quality of the fit to the molecular electrostatic potential in estimating the atomic point charges for each atom (figure 4.3).
The charge on the hydrogen in H-N (table 4.2) is bigger than that on the hydrogens in H-C. This is the most notable difference between the H atom (H1 in figure 4.3) involved in a hydrogen bond and the others, bonded to the C atoms. Both N atoms have a negative charge, while the H, involved in the intermolecular hydrogen bond H-N(1)···N(2), has positive charge, which is consistent with an electrostatic model of hydrogen bond and

also in good agreement with previous work about the charge density in imidazole[18]. The *ab-initio* calculated electron density distribution (figure 4.3) has an approximate symmetry *2mm*, which is the same observed at 103K and 293K[18], confirming that it does not change substantially with temperature. The fact that the symmetry of the electron density distribution around the molecule is the same at all temperatures, suggests that an eventual intermolecular proton transfer from N(1) to N(2), which might be possible at high temperatures, would not involve a major change in the imidazole charge distribution, within a possible protonic conduction mechanism in imidazole molecular chains[19]. We also calculated the molecular volume using the Cerius2[20] program, considering a fine grid spacing used to created a surface similar to a smooth van der Waals surface, in which we estimate the molecular volume. The van der Waals scale factor is 1.00 by default and the calculated molecular volume is 64.47Å^3.

4.3.2 Test of the FIT potential

Table 4.3 reports the results of the lattice energy minimisation both for the experimental and the optimised molecular geometry.

	Method (R-factor)	a /Å (δ%)	b /Å (δ%)	c /Å (δ%)	β /o	Unit cell Vol. /Å^3	Latt. Ener. /kJ mol^{-1}	*RMS %*	HB length /Å
Exp Struct (ref.)	Neutr. diff. (0.04)	7.569 -	5.366 -	9.785 -	119.1	347.32	-76.86	-	1.81
Exp Min ext. DMA	-	7.718 (1.96)	5.461 (1.78)	9.814 (0.30)	120.6	356.03	-78.38	1.54	1.83
Exp Min opt DMA	-	7.721 (2.00)	5.501 (2.52)	9.839 (0.56)	121.0	358.03	-75.14	1.88	1.80
Exp Min exp. CHelpG	-	7.773 (2.71)	5.301 (-1.22)	10.029 (2.49)	120.6	358.87	-61.88	2.06	2.10
Exp Min opt. CHelpG	-	7.739 (2.24)	5.350 (-0.29)	10.040 (2.60)	120.1	359.57	-60.25	1.99	2.08

Table 4.3: Values of the energies, lattice parameters and hydrogen bond lengths, obtained from an imidazole crystal lattice minimisation, with the use of multipole-expansion of *ab-initio* calculated atomic charges (*DMA*) and sole *ab-initio* calculated atomic point charges (*CHelpG*). *Exp-Min-exp* indicates the "Experimental minimised unit cell with experimental molecular geometry" and *Exp-Min-opt* indicates the "Experimental minimised unit cell with optimised molecular geometry". RMS coefficient calculated as: $RMS(\%) = \sqrt{\frac{(\delta a)^2 + (\delta b)^2 + (\delta c)^2}{3}}$ where δa (and similarly for δb and δc) has the expression: $\delta a = 100 \cdot \frac{(a_{exper.} - a_{calc.})}{a_{calc.}}$

The value of the lattice energy for the experimental molecular crystal structure accurately reproduces that obtained in previous calculations[5], using the same potential. The estimate of the lattice cell parameters is in good agreement with that previously obtained[5] as results from the RMS(%) value (figure 4.4). Following the minimization there is an increase of the unit cell volume with a difference of 3% compared to the experimental value.

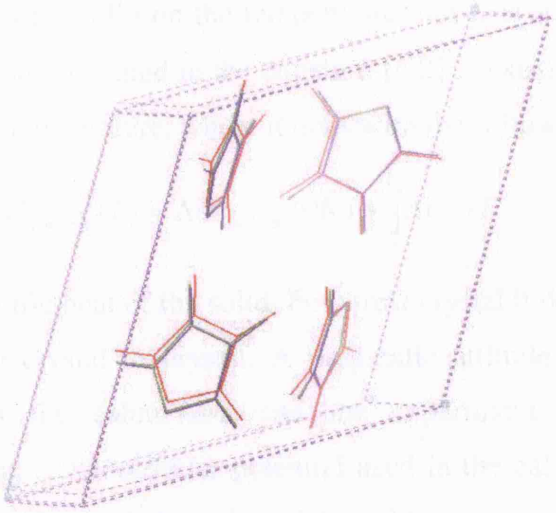


Fig. 4.4: Overlapped pictures of the three unit cells of the imidazole crystal, *Exp* (red), *Exp-Min-exp* (green), *Exp-Min-opt* (blue).

From table 4.3 we note that, for the two structures minimised with DMA of the atomic charges, there is a good agreement in the lattice energy with the experiment (*Exp Struct.* in table 4.3), while a larger discrepancy is observable for the two structures, minimised with the sole CHelpG atomic charges. The difference between the energies of the minimised structures with DMA in table 4.3 and sublimation enthalpy, which is 83 ± 0.2 kJ mol^{-1} (estimated through experimental measurements[21]) is due to the approximations in the model potential and in the neglecting of thermal effects in the calculations. As reported in the litterature[22], denoting as *P.P.E.* the calculated packing potential energy (which is half of the theoretical lattice energy U_{latt}), the sublimation enthalpy should be the same of the lattice energy of the crystal, as given by the formula [1.7] in chapter 1:

$$\Delta H_{subl.} \approx \frac{1}{2} P.P.E \approx U_{latt} \quad . \qquad [4.1]$$

This expression is only an approximation: in fact, while the *P.P.E.* does not refer to any temperature, the sublimation enthalpy is highly sensitive to the latter. Considering the equi-partition principle (for each degree of freedom of a molecule there is an energetic contribution of $1/2k_BT$, where k_B is the Boltzmann's constant), the thermal expansion of the solid with temperature and the molecular motion, we can better write (as previously reported in equation [1.11], chapter 1):

$$\Delta H_{subl.} = \Delta U_{condens} - 2RT \quad , \qquad [4.2]$$

where R is the ideal gas constant. Equation [4.2] shows the dependence of the sublimation enthalpy of a solid on the temperature, but it is still not completely correct. Another approximation contained in the equation [4.2] consists of considering the ΔU_{subl} not depending on the temperature, whilst it does with the following relationship:

$$\Delta U_{condens}(T) = \Delta U_{condens}(0\text{K}) + \int_0^T \Delta C_v dT \quad , \qquad [4.3]$$

where ΔC_v is the specific heat of the solid. For a real crystal it depends on the temperature and it changes from crystal to crystal. A pragmatic attitude is to consider that some differences between the calculation and the experiment are incorporated in the parametrisation of the intermolecular potential used in the calculation. In comparing the value of the lattice energy with the value of the sublimation enthalpy, a discrepancy of at most 8kJmol^{-1} is expected[23].

The lattice energy calculated following the minimisation of the optimized molecular crystal structure is different compared with that calculated for the experimental molecular geometry. This difference is small, bearing in mind that we neglected thermal effects. In particular, the lattice energy in the case of the optimized molecule is smaller than the value calculated for the experimental molecule, which can be explained by the planarity of the optimized molecule, which does not permit the slight bending of the N-H bond, involved in the intermolecular hydrogen bond interaction. This impediment does not facilitate the optimisation of the packing of the crystal and it also justifies the increase in the unit cell volume in the case of optimized molecular structure, with respect to that observed experimentally. The RMS(%) value for the unit cell parameters in the case of the crystal structure with *ab initio* optimised molecule is larger than that obtained for the crystal structure with the molecule in the experimental conformation, but it is still small enough to validate the simulations.

The results obtained from the lattice energy minimisation using only the atomic point charges without a multipole expansion are very different. As previously mentioned, the lattice energy is underestimated both for the structure with the experimental and the optimised molecular conformations, with a value between -60 and -62 kJ mol^{-1}. In particular ***c*** lattice parameter, along which is the intermolecular hydrogen bond, changes considerably. A reason of this discrepancy might be in the different positions of the N-H group in the two different molecular conformations (optimised and experimental), which is slightly out of ring plane in the experimental conformation with respect to the

optimised conformation (table 4.1), and affects the hydrogen bond distance, which takes place along the ***c*** axis of the unit cell.

4.3.3 Calculation of lattice phonon frequencies and elastic constants of experimental structure

Table 4.4 reports the frequencies of the lattice vibrations (phonons) calculated using DMA, either for the structure with the molecule in the experimental conformation (*ExpMinExp*), or for that with the molecule in optimised conformation (*ExpMinOpt*), which were calculated in the harmonic approximation, at $\boldsymbol{k} = 0$ (see appendix 3.2, chapter 3), using the DMAREL[14] software. The frequencies are contrasted with those measured at 100K (with Raman spectroscopy[24]) and 93K (with I.R. spectroscopy[25]) and with those calculated in a previous study, which performed MD simulations on the experimental minimised structure, at 103K[26].

		Exp. Struct.[24,25] (100K, 93K)	*ExpMinExp.* (MD-103K)[26]	*ExpMinExp.* (0K)	*ExpMinOpt.* (0K)
Mode	***Frequency Label***	***Phon. Freq.*** /cm^{-1}	***Phon. Freq.*** /cm^{-1}	***Phon. Freq.*** /cm^{-1}	***Phon. Freq.*** /cm^{-1}
A_g	ν_1	54.5	47	45.7	45.1
	ν_2	79.0	79	79.3	78.2
	ν_3	86.5	99	98.8	94.0
	ν_4	100.5	109-115	111.6	110.7
	ν_5	112.0	140	141.8	139.6
	ν_6	158.0	211	213.4	201.0
B_g	ν_7	62.5	56	54.8	54.2
	ν_8	74.0	88	86.5	83.3
	ν_9	109.0	106	108.0	105.2
	ν_{10}	128.0	154	157.0	150.0
	ν_{11}	163.0	181	184.1	177.2
	ν_{12}	181.0	206	206.2	195.6
A_u	ν_{13}	-	77	74.1	69.6
	ν_{14}	113.0	109-118	111.7	111.7
	ν_{15}	-	138	141.3	135.6
	ν_{16}	151.0	183	185.3	178.3
	ν_{17}	188.0	193	198.2	186.8
B_u	ν_{18}	70.0	63	61.4	61.4
	ν_{19}	97.0	102	102.3	98.0
	ν_{20}	128.0	147	149.9	148.2
	ν_{21}	-	204	209.2	196.1
RMS res.				***19.9***	***16.2***

Table 4.4: Imidazole phonon frequencies, experimentally measured at 100K and 93K, calculated from MD simulation at 103K[26] and calculated with DMAREL, within the harmonic approximation, for **k**=0.

For the calculated structure, with the molecule in the experimental conformation (*ExpMinExp*–0K in table 4.4), the frequencies are in acceptable agreement, with the values obtained from an MD simulation of the experimental structure at 103K[26] (*ExpMinExp*-MD-103K in table 4.4). Both of the calculated frequencies differ from the experimental values, measured at the temperature of 100K (with Raman Spectroscopy) and 93K (with IR)[5]. The normal modes ν_{11} and ν_{20} twist the molecules around the axis of the N-H···N bond, with minimal distortion of inter-chain contact. The larger values of the experimental frequencies with respect to those obtained from the calculations, might be due to the presence of anisotropy in the electrostatic interactions between the atoms[5].

For the calculated structure with the molecule in the *ab-initio* optimised conformation, the lattice vibrational frequencies are generally underestimated by up to 15 cm^{-1}, with respect to those calculated for the structures with the molecule in the experimental geometry, while for some modes the frequencies are quite similar. A possible reason for this general underestimation could be the constricted strong planarity of the molecule of the imidazole molecule, which does not favour the hydrogen bonding interactions and the vibrational modes of the lattice. However, comparing the RMS residual values, the harmonic approximation in the case of the optimised molecular structure (*RMS res* = 16.2) has a lower value than in the case of the experimental molecular structure (*RMS res* = 19.9). In the case of CHelpG calculation without the DMA, the RMS residual is definitely higher and the discrepancy in the frequencies values is larger than the discrepancy in the other calculations.

As we previously mentioned, the theoretical estimate of the elastic constants C_{ij} also involves the second derivatives of the potential energy surface at the classical structure, at 0K (appendix 3.2, chapter 3). In calculating these constants, zero-point energy and thermal effects are neglected. Table 4.5 reports the values of the diagonal constants C_{ij} for both the minimised crystal structure with the experimental molecular structure (*ExpMinExp*), and for the crystal structure based on the geometry optimised molecular conformation (*ExpMinOpt*).

	C_{11} /GPa	C_{22} /GPa	C_{33} /GPa	C_{44} /GPa	C_{55} /GPa	C_{66} /GPa
ExpMinExp *DMA*	10.40	10.26	77.11	9.45	1.60	4.10
ExpMinOpt *DMA*	10.06	10.00	71.89	9.34	1.50	3.94
ExpMinExp *ChelpG*	11.19	9.60	39.41	5.24	1.33	5.95
ExpMinOpt *ChelpG*	11.22	9.29	38.93	5.20	1.35	5.89

Table 4.5: Diagonal components of the elastic tensor for imidazole crystal, calculated at 0k in the harmonic approximation. *ExpMinExp* indicates the experimental minimised structure with the imidazole molecule in the experimental conformation. *ExpMinOpt* denotes the experimental minimised structure, with the molecule in the *ab initio* optimised conformation.

The elastic constants are the elements of the elastic stiffness tensor. The diagonal elements of the tensor (C_{ji}) describe the stiffness of the lattice to uniaxial compression (C_{11}, C_{22}, C_{33}) and to uniaxial shear (C_{44}, C_{55})[26].

In table 4.5 one of the three diagonal components related to the uniaxial compression (C_{33}) is much larger than the others. This anisotropy, which reflects the different lengths of the unit cell parameters, means that the crystal shows an anisotropic behaviour under compression, preferably oriented along the direction of the **c** lattice parameter. This could be related to the direction of the hydrogen bonds between the molecules, which are preferably oriented along this lattice parameter. The elastic constants for the crystal with the optimized molecular conformation (*ExpMinOpt*) differ by a small amount with respect to those based on the structure with the molecule in the experimental conformation (*ExpMinExp*), showing a small increase, which can be attributed to the rigidity and planarity present in the optimized geometry of the imidazole molecule. The values of the shear constants show a distinct strength of the lattice to shearing perpendicular to the **c** axis (C_{44}). The values of the elastic constants in the case of the minimisation using only ChelpG charges are completely different with respect to the elastic constants calculated with ChelpG + DMA and tend to be much lower.

4.3.4 Crystal Structure Prediction

We report in figure 4.5 the results obtained by the crystal structure prediction procedure, previously described, for the imidazole molecule.

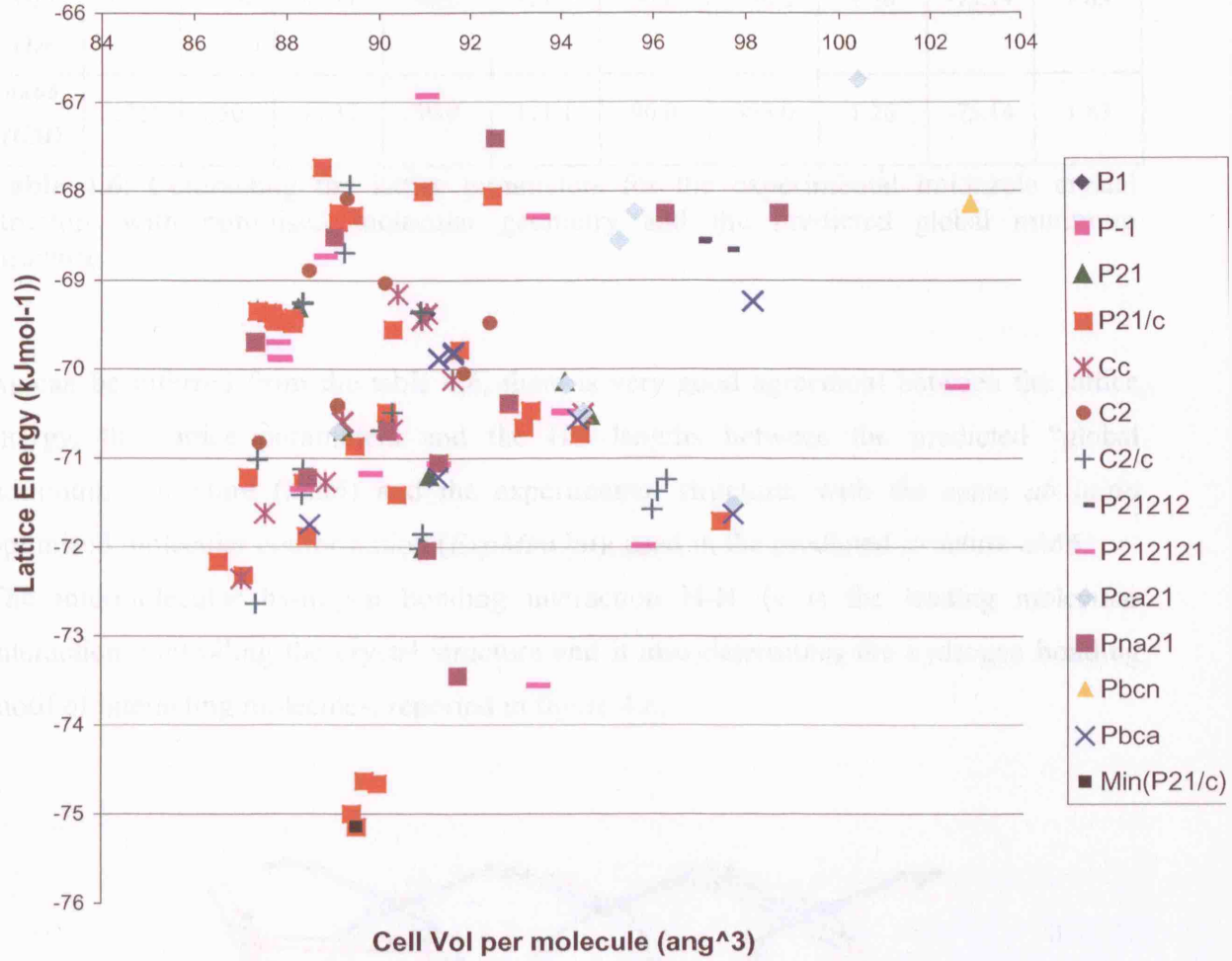


Fig. 4.5: Scatterplot of *lattice energy* vs *volume/molecule* in the imidazole crystal structure prediction.

The experimental structure of the imidazole, minimised with the molecule in the "gas phase" geometry was successfully predicted to be the global minimum (GM) in lattice energy, which corresponds to the predicted structure *ak66*, having space group $P2_1/c$ and lattice parameters reported in table 4.6.

	a /Å	b /Å	c /Å	α /°	β /°	γ /°	Vol /Å^3	ρ /gcm^{-3}	Latt. En. /kJ mol^{-1}	HB length /Å
Exp Min Opt	7.721	5.501	9.839	90.0	121.0	90.0	358.0	1.26	-75.14	1.83
ak66 (GM)	7.721	5.501	9.839	90.0	121.1	90.0	358.0	1.26	-75.14	1.83

Table 4.6: Contrasting the lattice parameters for the experimental imidazole crystal structure with optimised molecular geometry and the predicted global minimum structure.

As can be inferred from the table 4.6, there is very good agreement between the lattice energy, the lattice parameters and the HB lengths between the predicted "global minimum" structure (*ak66*) and the experimental structure, with the same *ab initio* optimised molecular conformation (*ExpMinOpt*), used in the predicted structure *ak66*.
The intermolecular hydrogen bonding interaction H-N···N is the leading molecular interaction controlling the crystal structure and it also determines the hydrogen bonding motif of interacting molecules, reported in figure 4.6.

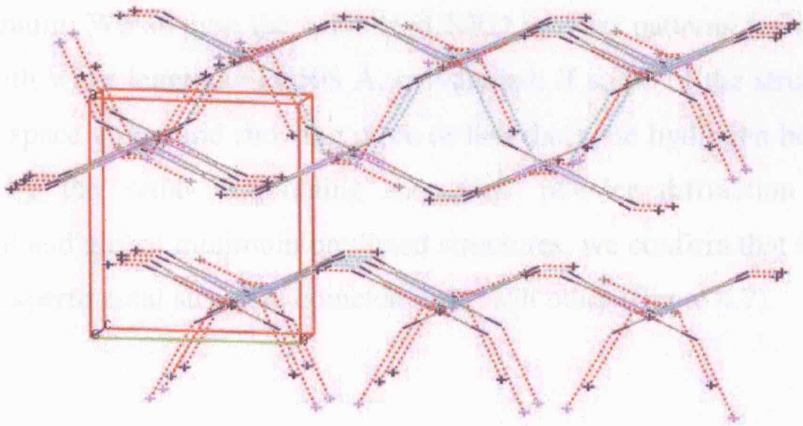


Fig. 4.6: Imidazole hydrogen bonding motif for the lowest lattice energy predicted crystal structure with (*ak 66-GM*), corresponding to the *ExpMinOpt* structure.

It consists of chains of hydrogen bonding interacting molecules along the **c** axis. Along each chain the molecules are alternately twisted with respect to each other with a torsion

angle of 52.46°. This kind of motif is in agreement with the type identified in previous work on the crystal structure of imidazole and imidazole derivatives[27], which report a torsion angle of 57° ~ 60°, in a same chain motif. The stacking interactions of these chains propagate along a plane orthogonal to the chain direction. Adjacent chains interact with each other, forming a sort of layer. The layers interact by short contacts, in pairs. The closest contact between adjacent chains measures 2.30 Å, whilst the closest contact between two layers is 2.71 Å, with a (C)H···N interaction. This kind of structure does not tolerate any substituent larger than hydrogen in the imidazole molecular ring, so that the resulting chain stacking is unique for the unsubstituted imidazole.

The N···N intermolecular distance in the hydrogen bonding interaction is 2.83 Å, smaller than the same distance for the crystal structure of imidazole derivatives (2.89 Å for mono-substituted, 2.90 Å for di-substituted and 2.94 Å for tri-substituted), reported in previous work[27]. This observation is sensible, considering the role of the steric interactions, which increase the distance N···N. The HB angle (H-N···N angle) is 168.85°, in good agreement with the value of ~162°, reported in a previous study[27].

It is worth highlighting that the H-N···N interactions are orientated with the reverse direction between two chains, differing from each other by 180°, resulting a *non-polar* crystal structure.

In the scatter-plot of the lattice energy *vs* unit cell volume per molecule there is a clear presence of different P21/c structures in a very small range of 0.5kJ/mol close to the global minimum. We analyse the calculated XRD powder patterns (using the Mercury[28] program) with wave length λ=1.0508 Å, to establish if some of the structures belonging to the same space group and showing more or less the same hydrogen bonding motif can be essentially the same. Examining the XRD powder diffraction pattern of the experimental and global minimum predicted structures, we confirm that the predicted and minimised experimental structure coincide with each other (figure 4.7).

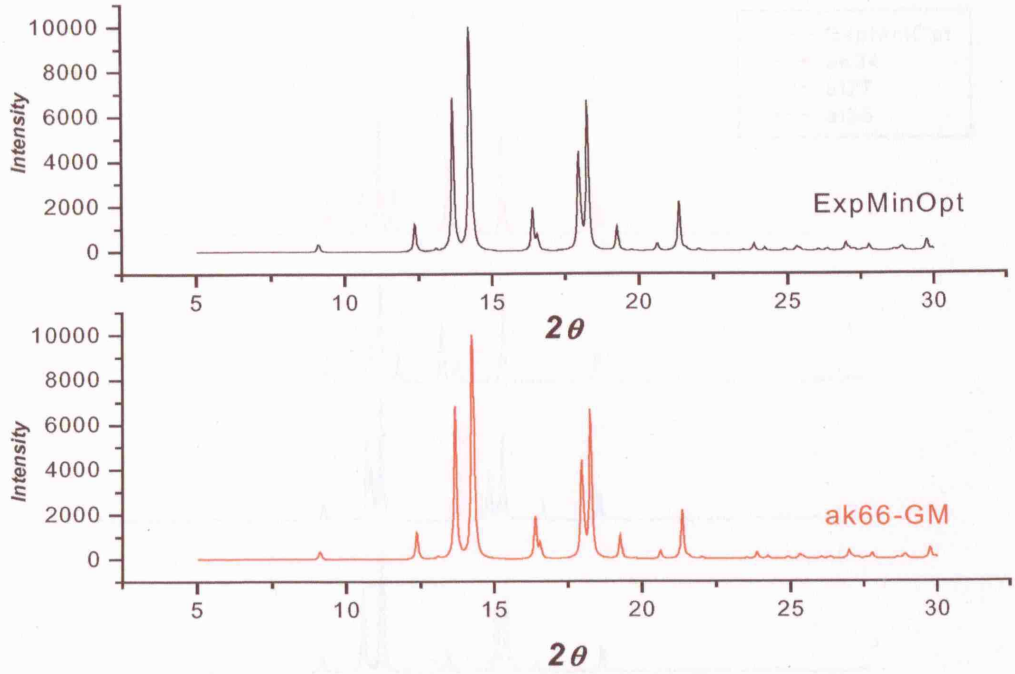


Fig. 4.7: XRD spectra (λ=1.0508 Å) of the experimental minimised structure (black) and the predicted ak66 global minimum structure (red), for the imidazole molecule.

Examining the XRD calculated patterns for the other structures in the $P2_1/c$ space group we can observe slight differences with respect to the diffraction pattern of the experimental minimised structure (see figure 4.8)

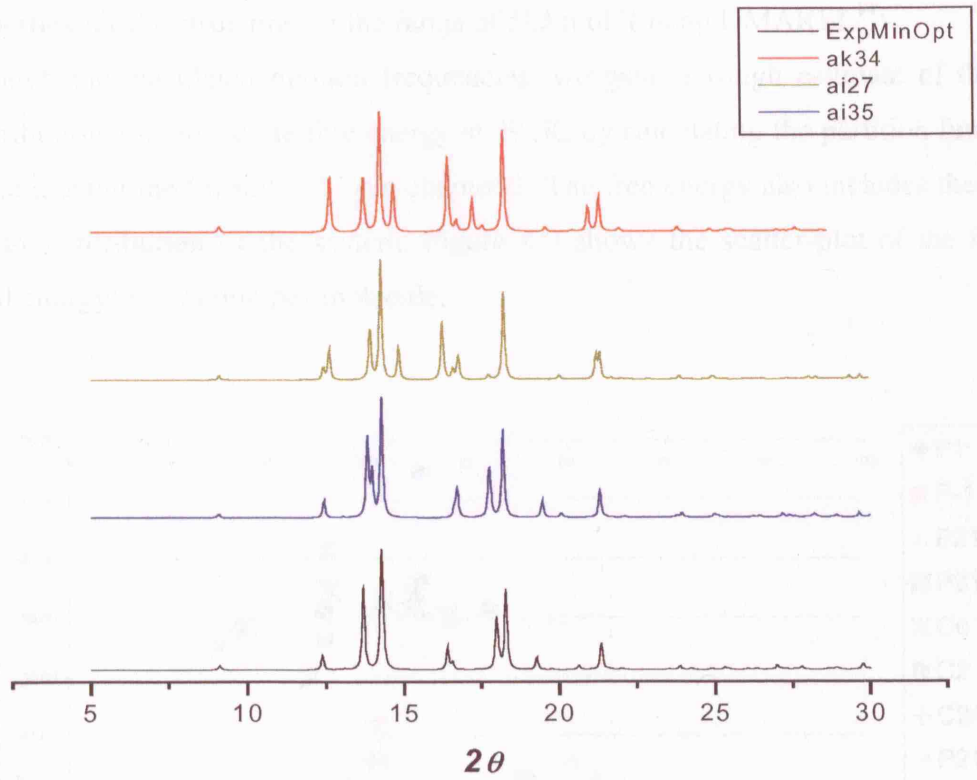


Fig. 4.8: XRD spectra of different predicted crystal structures of imidazole, in space group $P2_1/c$. The wave length of the radiation is λ=1.0508 Å .

The patterns show a similar structure, but with some differences and the appearance of new "shoulders" and peaks in *ak61* and *ak34* structures, which is quite surprising, considering that the difference in energy among these structures is less than the thermal energy kT (k is the Boltzmann's constant, T is the temperature in K). A sensible explanation is that from a static point of view the structures are "apparently" different, but considering the effect of the thermal motion and the phonon vibrations contributions, these differences could disappear and all of the structures end up in the "global minimum". Attempts to study the effect of thermal motion on the systems have been performed by the use of MD simulation and are discussed in section 4.4.

The problem often incurring in thermodynamic studies of the organic solid state is the presence of many different structures that are thermodynamically stable in a very small range of lattice energy[3], usually less than 10kJ mol^{-1}. The same problem is found in the present study, with the presence of many energetically stable structures with different space groups in a range of 5 kJ mol^{-1}.

In attempting to make more accurate predictions, we calculated thermodynamic properties for the structures in the range of 5kJ mol^{-1}(using DMAREL[14]).

Through the calculated phonon frequencies, we gave a rough estimate of the entropic contribution and hence the free energy at 298K, by calculating the partition function for $\boldsymbol{k} = 0$ and using the formula [2.7] in chapter 2. The free energy also includes the zero point energy contribution of the system. Figure 4.9 shows the scatter-plot of the free energy (total energy) *vs* volume per molecule.

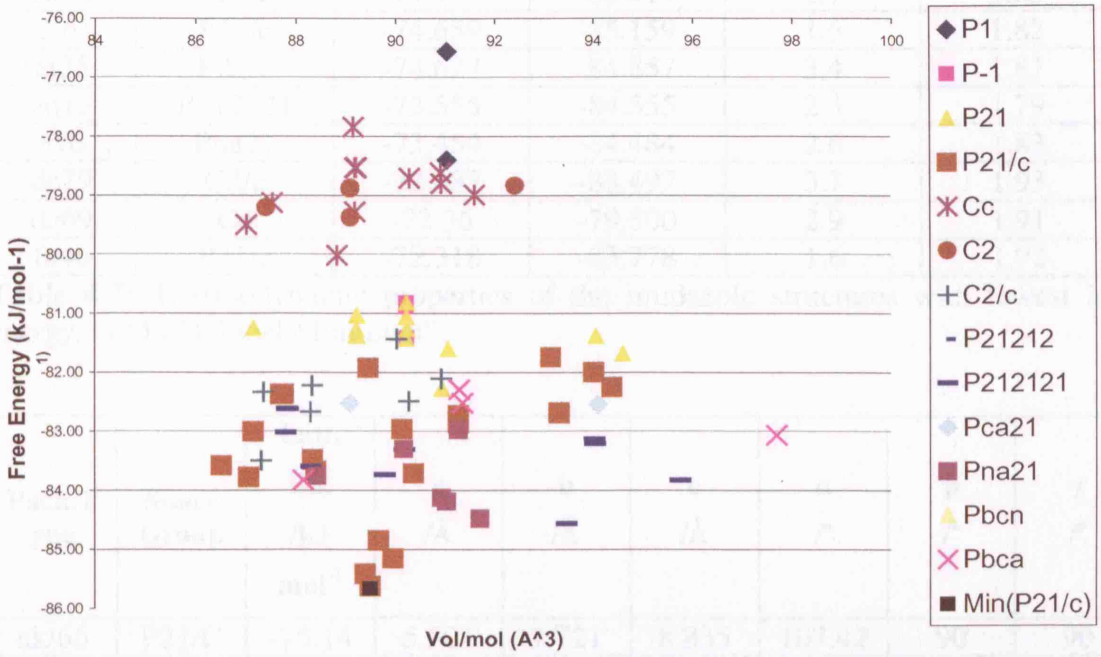


Fig. 4.9: Scatterplot of *free energy* vs *volume/molecule* for imidazole crystal structures prediction.

It is interesting to note that the two space groups $P2_12_12_1$ and $P2_1$, mixed among them in the lattice energy scatter-plot, are now distinct, with major stability for the structures in the $P2_12_12_1$ space group.

The calculation of the free energy confirms the higher stability of the structures with $P2_1/c$, $Pna2_1$ and $P2_12_12_1$ space groups with respect to the others. Having to choose among many possible structures, we focused our attention on the properties and the hydrogen bonding motif for the structures in the lattice energy range between 75 and 72 $kJmol^{-1}$.

The thermodynamic properties of these structures are summarised in tables 4.7 and 4.8, reporting the lattice energy, the free energy, the lowest shear constant of the elastic tensor

(which is related to the mechanical stability of the structure) and the hydrogen bonding length.

Packing Type	Space Group	Lattice En. /kJ mol^{-1}	Free En. /kJ mol^{-1}	Lowest Shear Const. /Gpa	HB lengths /Å
ExptMin Opt	*P21/c*	*-75.138*	*-85.668*	*1.5*	*1.83*
ak66 (GM)	P21/c	-75.143	-85.623	3.7	1.83
ak34	P21/c	-74.997	-85.417	1.7	1.83
ak61	P21/c	-74.659	-85.159	1.6	1.82
ai35	P21/c	-74.627	-84.857	3.4	1.81
aq13	P212121	-73.555	-84.555	2.3	1.79
av6	Pna21	-73.454	-84.484	2.6	1.83
dc79	C2/c	-72.637	-83.497	3.7	1.93
da69	Cc	-72.36	-79.500	2.9	1.91
fa46	P21/c	-72.318	-83.778	1.6	1.92

Table 4.7: Thermodynamic properties of the imidazole structures with lowest lattice energy. G.M.="Global Minimum".

Pack.Type	Space Group	Latt. En. /kJ mol^{-1}	a /Å	b /Å	c /Å	α /°	β /°	γ /°
ak/66	P21/C	-75.14	5.501	7.721	8.835	107.42	90	90
ak/34	P21/C	-74.99	5.502	6.81	9.823	103.69	90	90
ak/61	P21/C	-74.66	5.492	6.749	9.893	101.1	90	90
ai/35	P21/C	-74.63	5.474	7.591	9.076	108	90	90
aq/13	P2121 21	-73.56	5.423	7.177	9.605	90	90	90
av/6	Pna21	-73.45	4.738	8.47	9.141	90	90	90
dc/79	C2/c	-72.64	4.906	8.912	16.436	103.59	90	90
da/69	Cc	-72.36	5.32	8.021	8.641	90	90	109.28
fa/46	P21/C	-72.32	5.048	5.081	15.153	90	90	116.35

Table 4.8: Lattice energy, cell parameters and unit cell densities of the imidazole structures with lowest lattice energy. GM="Global Minimum".

The values of the elastic shear elastic constants show a good mechanical stability for all structures and the rank of stability is the same for the five lowest energy minima.

The hydrogen bonding motifs of the *ak61* and *ai35* structures are similar to each other, although there is a slight difference between their powder patterns; the structures consist

of the same interacting chains of hydrogen bonding alternately twisted molecules along the **c** axis as observed for the global minimum structure. The difference with respect to the "global minimum" structure is that there is no (C)H···N interaction between the layers, which causes the larger distance between the groups involved. Another possible structure with space groups P21/c is *fa46*, which has a flat tapes structure. Two chains interact in a "herring bone" pattern as shown in figure 4.10:

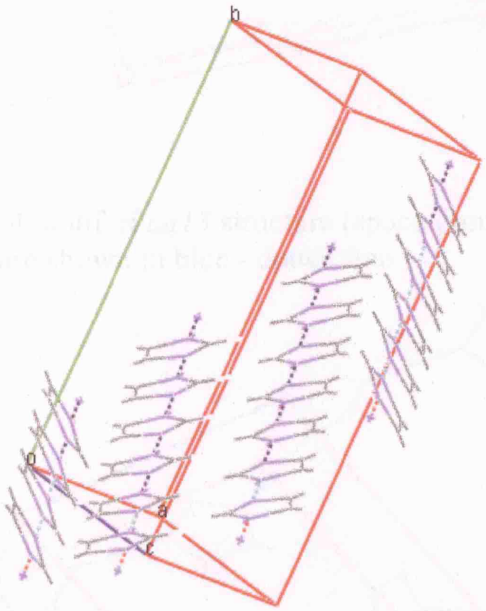


Fig. 4.10: Alternative low energy hydrogen bond motif (*fa46*) for an imidazole crystal in P21/c space group. The hydrogen bonds are shown in blue - dotted line

In figures 4.11- 4.14 are shown the hydrogen bond motifs of some of the other low energy structures in space groups different than P21/c, which are reported in the table 4.7.

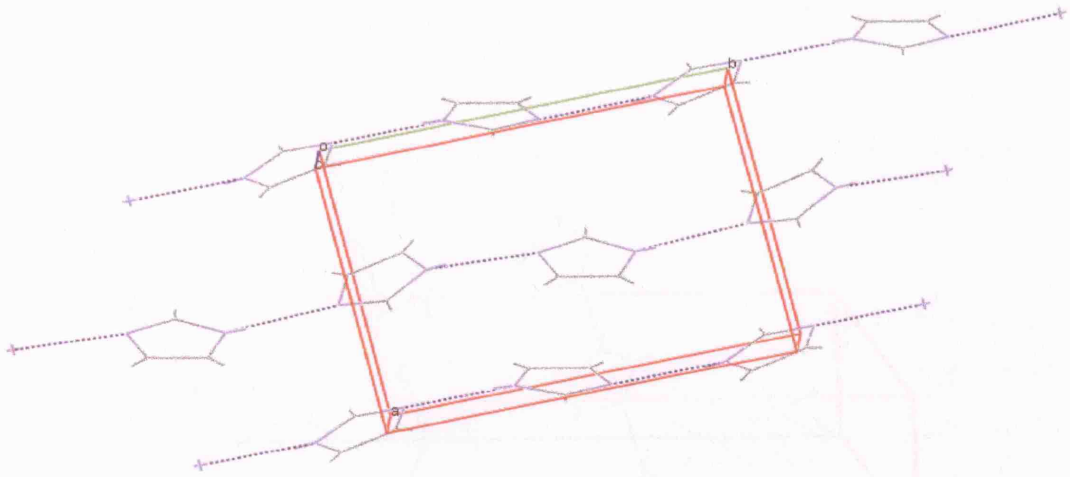


Fig. 4.11: Hydrogen bond motif of *aq13* structure (space group $P2_12_12_1$) in "*wavy*" shape chains. Hydrogen bonds are shown in blue - dotted line.

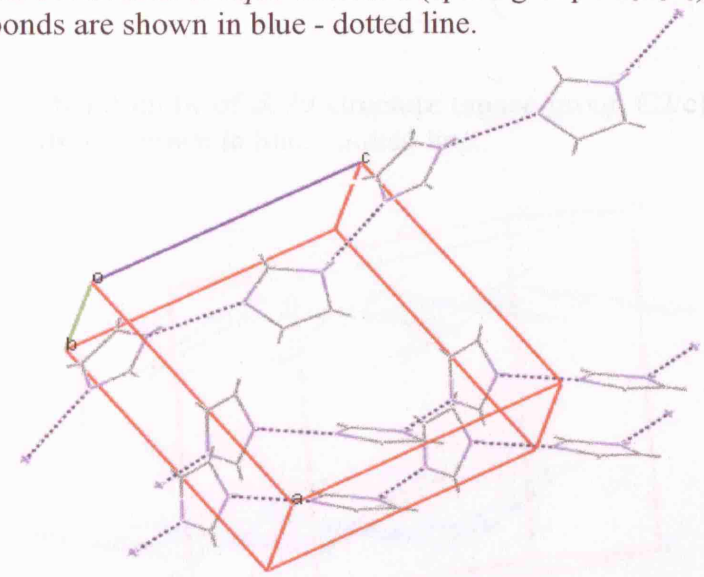


Fig. 4.12: Hydrogen bond motif of *av6* structure (space group Pna2). Hydrogen bonds are shown in blue - dotted line.

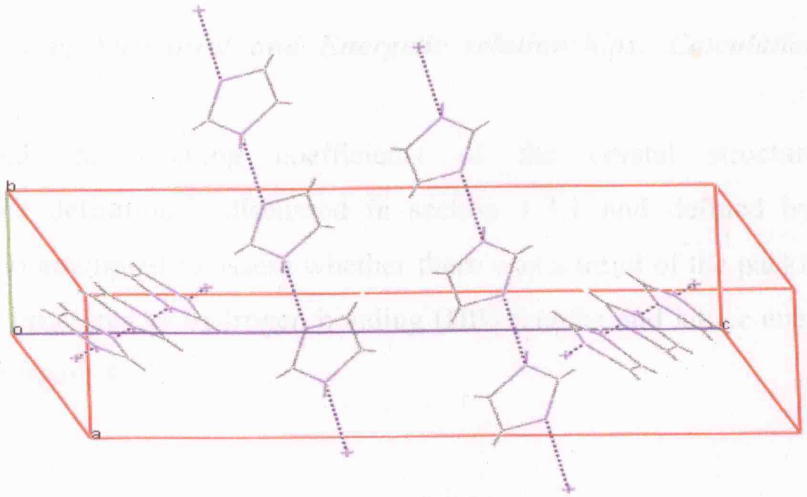


Fig. 4.13: Hydrogen bond motif of *dc79* structure (space group C2/c) in *planar* shape chains. Hydrogen bonds are shown in blue - dotted line.

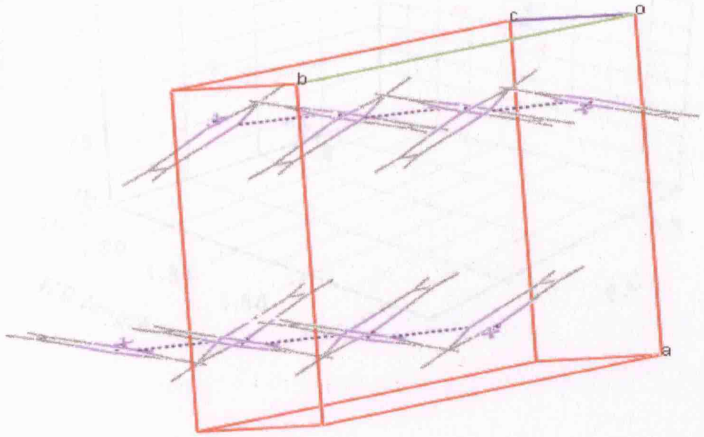


Fig. 4.14: Hydrogen motif of the close local minimum structure *ai35* ($P2_1/c$), in chains with twisted adjacent molecules. Hydrogen bonds are shown in blue - dotted line.

It is interesting to note the presence of the planar chains in most of these structures, apart from the $P2_12_12_1$, which shows a wavy sheet parallel to the ***b*** axis of the unit cell. Although there are different spatial orientations and shapes of the chains ("*twisted* ", "*wavy*", "*planar*") for different crystal structures, the hydrogen bond motif in all of these structures is always the same and is characterised by the N-H···N intermolecular interaction. The transition from a structure of a "*wavy*" chain to a lower energy structure, with a "*twisted*" chain, would require to the system to overcome only a low potential energy barrier. This factor, together with the similarity of the hydrogen bond motifs

among the lowest energy structures, suggests that metastable structures are unlikely to be effective new polymorphs of the imidazole molecule.

4.3.5 Establishing Structural and Energetic relationships: Calculation of Packing coefficients

We calculated the packing coefficients of the crystal structures, following Kitaigorodsky's definition[29], discussed in section 1.3.1 and defined by the equation [1.17]. We also attempted to assess whether there was a trend of the packing coefficients of the crystal structures *vs* hydrogen bonding (HB) lengths and lattice energy, using 3D-plot, shown in figure 4.15.

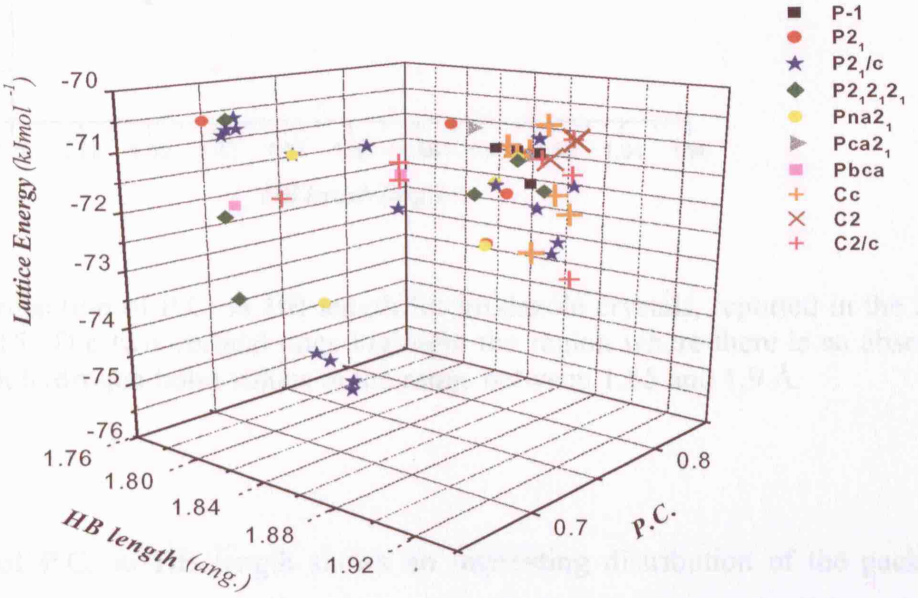


Fig. 4.15: 3D-plot lattice energy *vs* P.C. *vs* HB length for imidazole crystals.

The typical values of the packing coefficient for organic molecular crystals ranges from 0.63 to 0.78 as reported in the literature[16]. Figures 4.16 and 4.17 show the different 2D-projections of the 3D-plot, reported in figure 4.15

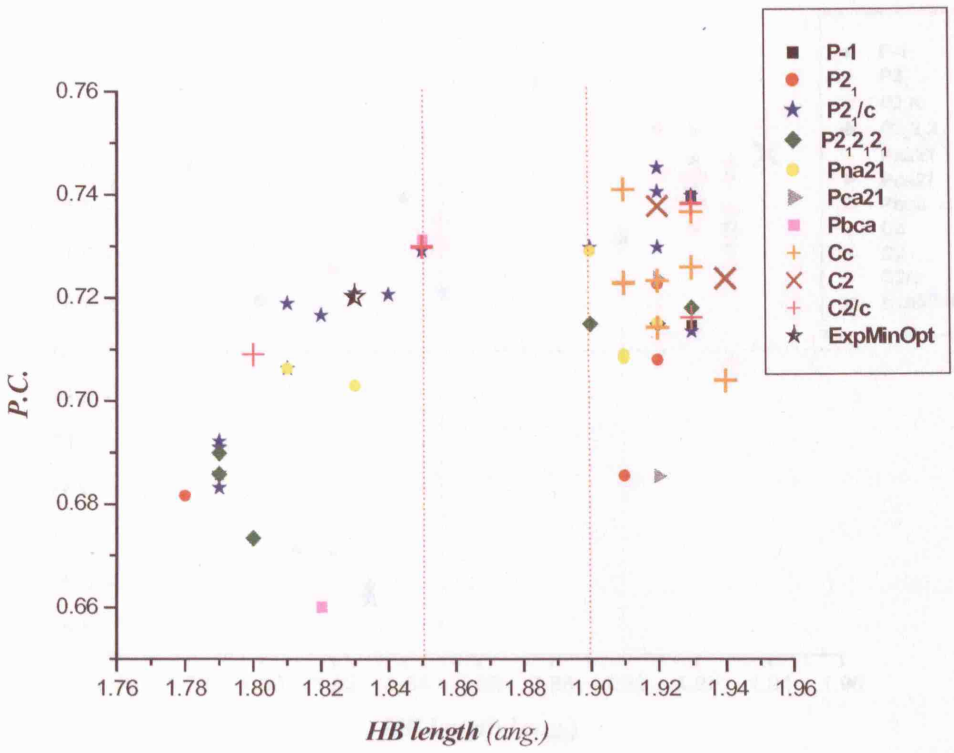


Fig. 4.16: 2D projection of P.C. *vs* HB length for imidazole crystals, reported in the 3D-plot in figure 4.15. The two vertical lines highlight the region where there is an absence of structures with hydrogen bond values in the range between 1.85 and 1.9 Å.

The projection of P.C. *vs* HB length shows an interesting distribution of the packing coefficients, which are not directly related to the HB length in the intermolecular interactions. Both the lowest energy and the experimental structures do not have the highest value of the P.C. and in general it is not true that the larger the HB length, the lower is the crystal density. In different space groups, a large value of P.C. is associated with a large value of the HB length (~ 1.9Å), which in general implies an increase in the lattice energy as we observe in the figure 4.17, in the group of points in the region at the right of the second vertical dotted line. For some space groups (i.e. $P2_1/c$), we can note a decrease of the packing coefficient for short HB lengths (figure 4.16), with a corresponding increase in the lattice energy (figure 4.17). The values of the packing coefficient corresponding to the lowest lattice energy structures are around 0.71 and 0.72 and correspond to an HB length in the range between 1.81 and 1.83Å.

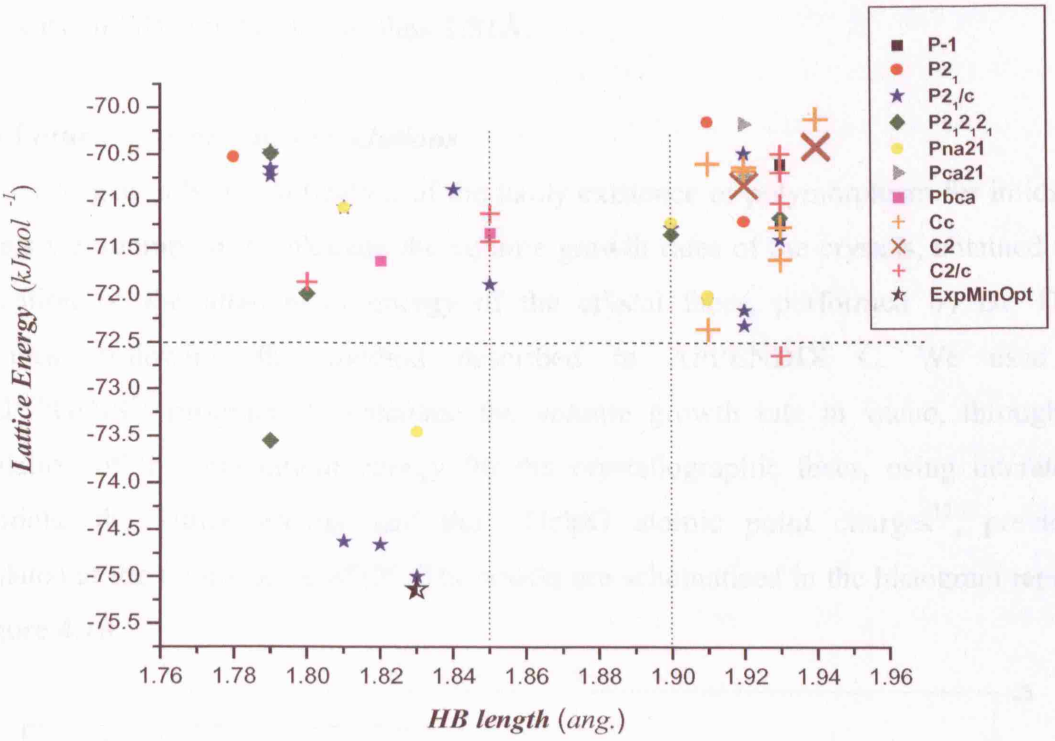


Fig. 4.17: Lattice energy *vs* HB length distribution for imidazole crystals. The two vertical lines highlight the region where there is an absence of structures with hydrogen bond values in the range between 1.85 and 1.9Å.

Both the two 2D-projections of the 3D-plot show a range of HB lengths in two different intervals. The HB distances are shorter than 1.86 or longer than 1.90. There are no structures with an HB distance between these limit values. The structures having lattice energy in the range of 3kJmol^{-1} above the global minimum have mainly HB lengths shorter than 1.86, apart one structure in the space group C2/c, which has HB length ~ 1.93 Å. The structures belonging to the space group Cc seem to have long HB lengths and simultaneously large P.C. These characteristics explain the generally high lattice energy of these structures, with respect to the global minimum. To summarise we can note some general correlations between lattice energy, P.C. and hydrogen bond lengths. The denser the structures the more stable they are, while the shorter the interacting distances between atoms of different molecules, the denser and more stable is the crystal structure. This trend is not observed in the case of hydrogen bonds, in particular in imidazole structures with very short HB lengths, which involve a high repulsive wall

between the nitrogen atoms belonging two different molecules, which point to each other in the case of HB lengths shorter than 1.81Å.

4.3.6 Volume growth rate calculations

In order to gain a better indication of the likely existence of polymorphism for imidazole crystals we attempted to calculate the volume growth rates of the crystals, obtained from calculation of the attachment energy of the crystal faces, performed by Dr. David Coombes, following the method described in APPENDIX C. We used the GULP[30]/GDIS[31] program, to calculate the volume growth rate in vacuo, through the calculation of the attachment energy for the crystallographic faces, using interatomic potentials, the lattice energy and the CHelpG atomic point charges[12], previously calculated at the temperature of 0K. The results are schematised in the histogram reported in figure 4.18.

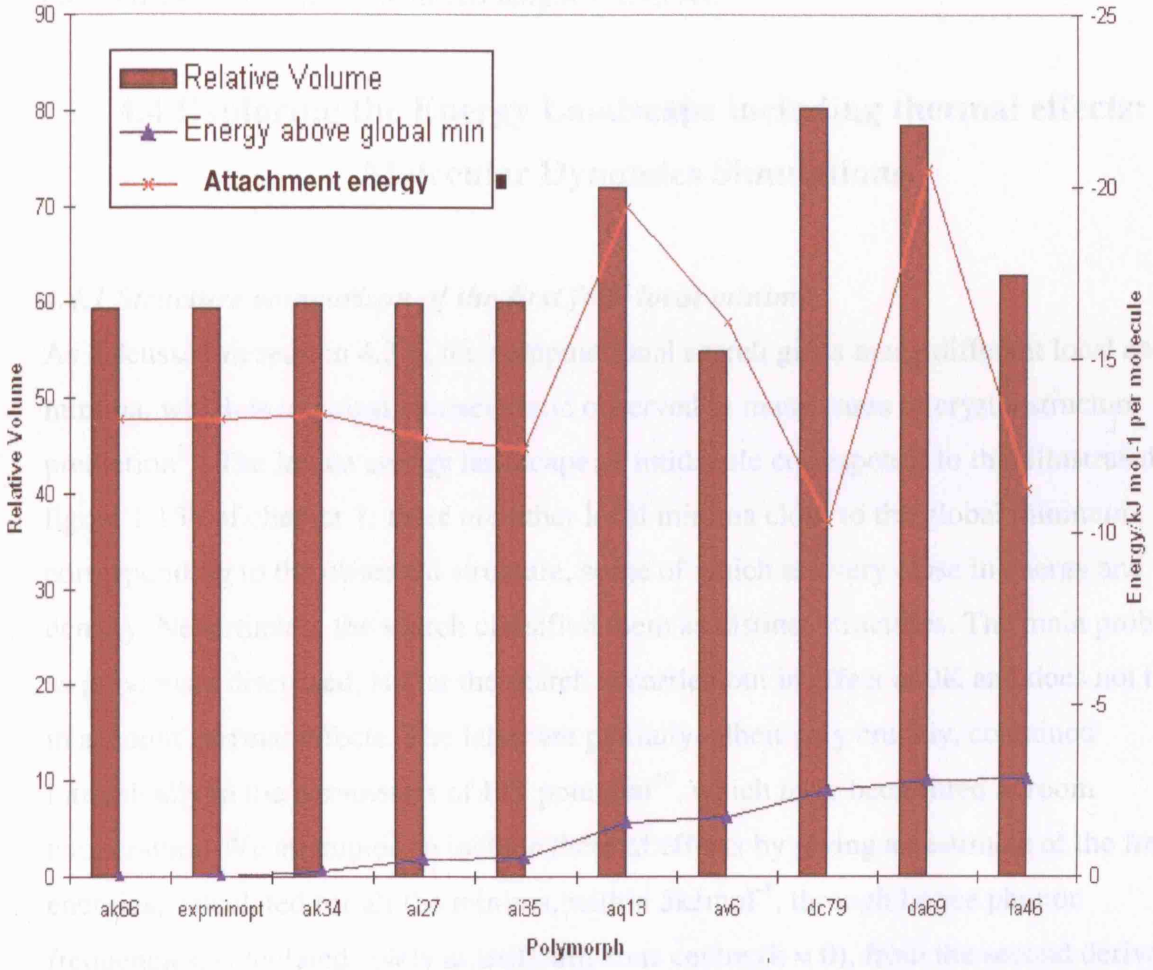


Fig. 4.18: Relative volume growth rate and attachment energy calculation for imidazole molecular crystals.

The volume growth rate is essentially the same for the structures in the P21/c space group, which are very close in energy and seem to have the same growth kinetics. The structure with the *av6* packing type has a lower growth rate and considering its complicated packing motif, it is probably not a likely candidate for an observable polymorph. Another structure (*da69*) with space group Cc has a high growth rate volume. This structure has polarized surfaces, which usually lead to the formation of twinned crystals, or strongly solvated crystals in order to annul the dipole. Moreover the presence of this polarized surface can invalidate the estimate of the relative volume growth rate of the structure, which reduces confidence in our results. The other two structures *aq19* (space group $P2_12_12_1$) and *dc79* (space group C2/c) have a high volume growth rate, larger than the experimental structure; they also have good elastic constants and free energy values indicting both mechanical and thermodynamic stability. The packing coefficients are P.C. ~ 0.69 for the *aq13* packing type (with a HB length ~ 1.79 Å) and P.C. ~ 0.74 for the *dc79* (with HB length ~ 1.93 Å).

4.4 Exploring the Energy Landscape including thermal effects: Molecular Dynamics Simulations

4.4.1 Structure comparison of the first four local minima

As discussed in section 4.3.4, the computational search gives many different local energy minima, which is a typical characteristic observed in many cases of crystal structure prediction[32]. The lattice energy landscape of imidazole corresponds to that illustrated in figure 1.15b of chapter 1: there are other local minima close to the global minimum, corresponding to the observed structure, some of which are very close in energy and density. Nevertheless the search classified them as distinct structures. The main problem, as previously discussed, is that the search is carried out in effect at 0K and does not take in account thermal effects. The latter are partially, albeit very crudely, contained intrinsically in the parameters of FIT potential[10], which have been fitted at room temperature. We attempted to include thermal effects by giving an estimate of the free energies, calculated for all the minima, within 5kJmol^{-1}, through lattice phonon frequencies, calculated solely at Brillouin zone centre (k = 0), from the second derivative of the potential with respect to the atom positions. The result was not encouraging as shown in figure 4.9, because the ranking of the structures was not substantially modified, especially for the lowest energy structures.

An alternative approach, which allows us to investigate whether structures are inter-convertible, is to use molecular dynamics (MD) simulation[23,26,33-35]. The technique, discussed in chapter 3, can be used also to calculate phonon frequencies of a certain crystal structure, but the method is very expensive in computer time[35]. We consider, on the basis of a previous work[36], that the technique is suitable for simulating room temperature transitions of structures located on the lowest energy minima to the global minimum structure, by giving an explicit description of the effect of the temperature on the system[36]. In particular if energy barriers among the local minima are of the order of thermal energy, we should observe a transition from the local to the global minimum structure. Figure 4.19 shows the labels of the atoms in imidazole molecule used in the MD simulation, which are used in the calculation of the radial distribution function (RDF) at different temperatures, illustrated in figures 4.39-4.41.

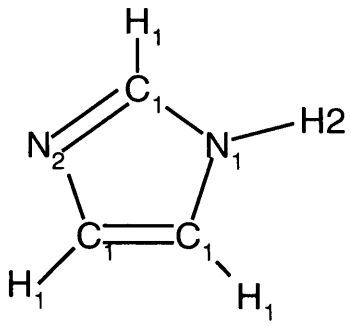


Fig. 4.19: Labels of the atomic species in the imidazole molecule, used in the MD simulations, performed with the program DL_POLY[37].

We initially performed a graphical comparison (with the Mercury[28] program) of all the local minimum structures, close in lattice energy to the global minimum structure. The comparison of all these structures is illustrated in figures 4.20- 4.23.

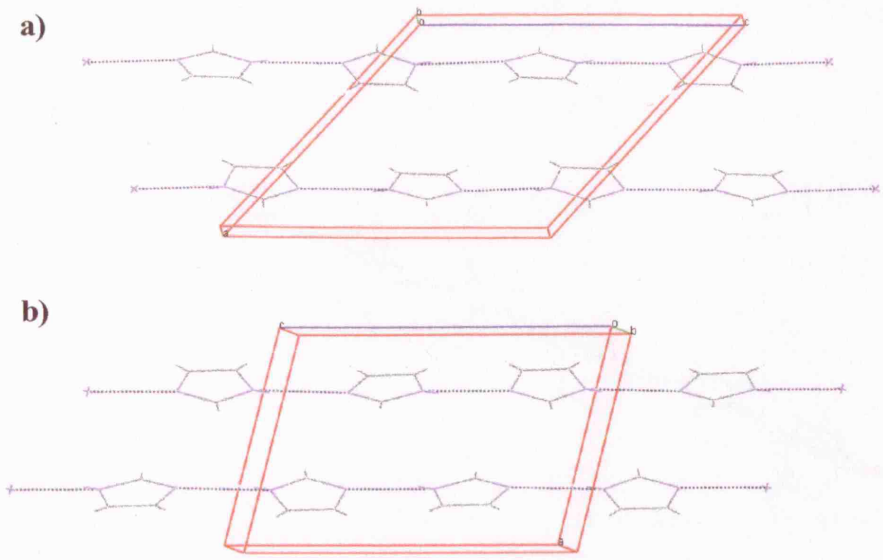


Fig. 4.20: Graphical comparison between the **a)** predicted "global minimum" *ak66*-GM and **b)** the local minimum *ak34* structures.

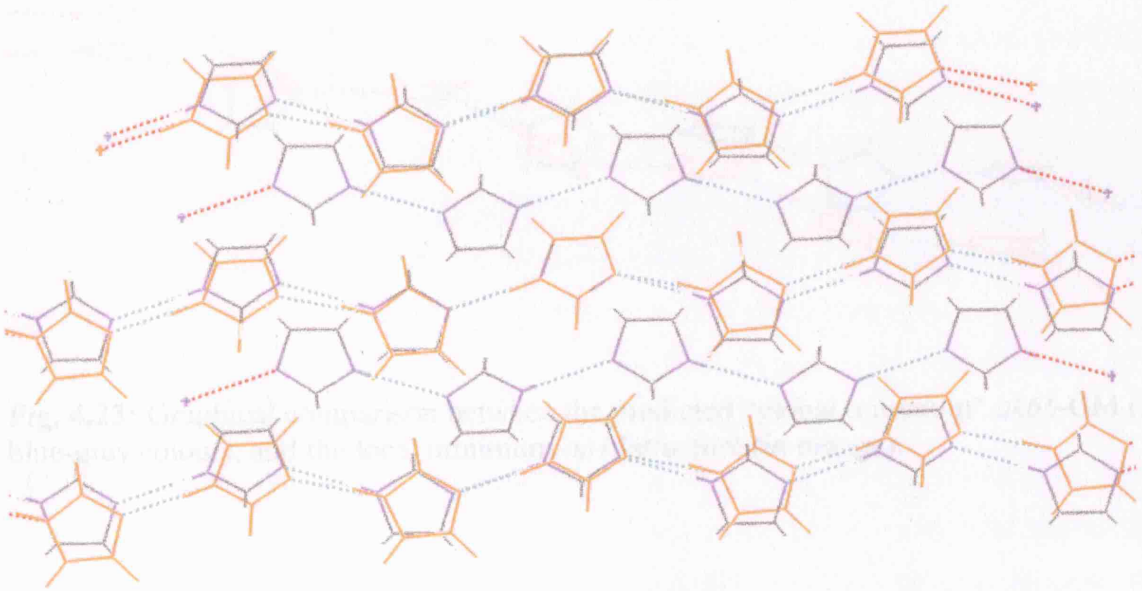


Fig. 4.21: Graphical comparison between the predicted "global minimum" *ak66*-GM (in blue-grey colour), and the local minimum *ak61* structure (in orange).

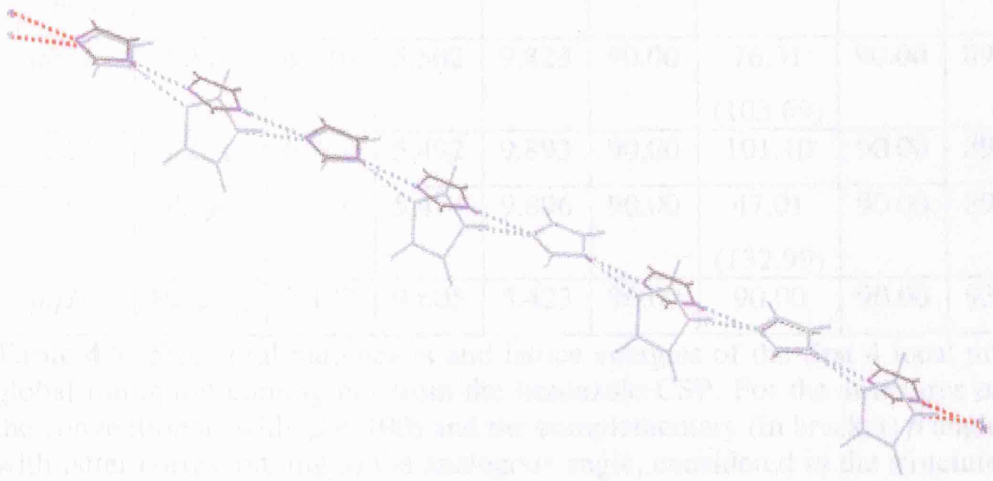


Fig. 4.22: Graphical comparison between the predicted "global minimum" *ak66*-GM (in blue-grey colour), and the local minimum *ai35* structure (in light blue).

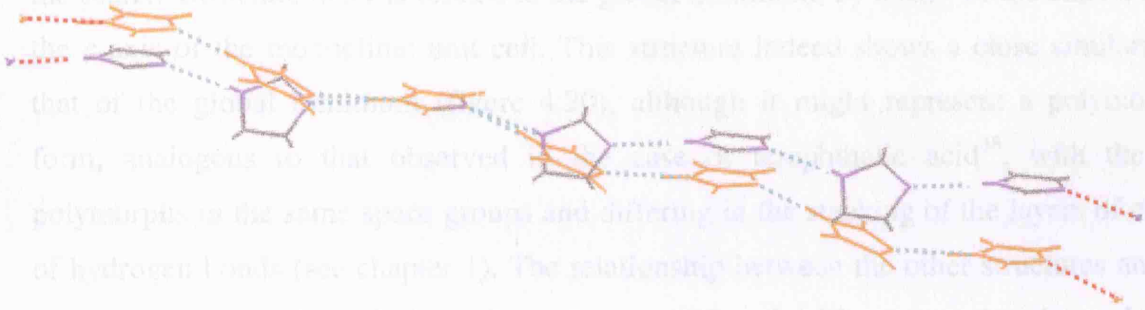


Fig. 4.23: Graphical comparison between the predicted "global minimum" *ak66*-GM (in blue-grey colour), and the local minimum *aq13*structure (in orange).

The structural parameters and energies corresponding to these local minima are reported in the table 4.11:

Packing type	*Space Group*	*a (Å)*	*b (Å)*	*c (Å)*	*α (°)*	*β (°)*	*γ (°)*	*Vol/Z (Å^3)*	*En (kJ mol^{-1})*
ak66 (GM)	$P2_1/c$	7.721	5.501	9.839	90.00	121.05	90.00	89.51	-75.14
ak34	$P2_1/c$	6.810	5.502	9.823	90.00	76.31 (103.69)	90.00	89.96	-74.99
ak61	$P2_1/c$	6.749	5.492	9.893	90.00	101.10	90.00	89.41	-74.66
ai35	$P2_1/c$	9.076	5.474	9.896	90.00	47.01 (132.99)	90.00	89.67	-74.62
aq19	$P2_12_12_1$	7.177	9.605	5.423	90.00	90.00	90.00	93.46	73.56

Table 4.9: Structural parameters and lattice energies of the first 4 local minima and the global minimum coming out from the imidazole CSP. For the structures *ak34* and *ai35*, the conventional (with $\beta < 100$) and the complementary (in bracket) β angle are reported, with latter corresponding to the analogous angle, considered in the structures of the other minima.

From figures 4.20-4.23, we observe certain analogies between the local minima and the experimental structure, which corresponds to the global minimum structure predicted in the search. Structure *ak34* is related to the global minimum, by a shift of the chains along the ***c*** axis of the monoclinic unit cell. This structure indeed shows a close similarity to that of the global minimum (figure 4.20), although it might represent a polymorphic form, analogous to that observed in the case of terephthalic acid[38], with the two polymorphs in the same space groups and differing in the stacking of the layers of chains of hydrogen bonds (see chapter 1). The relationship between the other structures and the global minimum is less obvious. From figures 4.22 and 4.23 we note that the molecules along the chains of the structures *ai35* and *aq13* are rotated with respect to the relative position of the molecules in the chain of the global minimum *ak66-GM* structure, although the two predicted local minima are similar to the global minimum structure and have similar hydrogen bonding motif. Some rotations seem to involve large angles like in the case of *ai35*. The structure of *ak61*, on the contrary seems to be more closely related to the structure of *ak66-GM*, with a likely mechanism of shift of the position of the layers with respect to those in the *ak66-GM* structure (figure 4.21). In order to have a more quantitative comparison of these structures, we used the program COMPACK[39], which performs a comparison of two crystal structures, by evaluating the root mean square (r.m.s) or discrepancy in the atomic positions between two overlapped clusters of

molecules, which are extracted from the two structures under contrast, and in which one of the molecules is chosen as reference point for the comparison of the two clusters. The standard clusters size is 15 molecules, and the tolerance in the distance between two corresponding atomic species can be varied until the r.m.s. value can be calculated by the program. In table 4.10 we report the r.m.s. values calculated for the four structures, corresponding to the local minima, with respect to the global minimum structure.

ak66-GM	*Tolerance*	*r.m.s (Å)*
ak34	50%	0.893
ak61	135%	2.008
ai35	135%	1.322
aq13	100%	1.769

Table 4.10: Root mean squares (r.m.s) values of the discrepancy in the atomic positions of four local minima structures with respect to the global minimum one (*ak66-GM*). The tolerance is expressed as percentage of atomic distance.

As we can see from the table, apart in the case of *ak34*, for all the other structures it was necessary to use a very big tolerance in matching the atomic positions of the structures, in order to have some correspondence. The values of the r.m.s. are very large in the case of *ak61*, *ai35* and *aq13* structures. It is worthy to note that a big value of r.m.s. is found even for the more similar structure *ak34* (~ 0.9 Å). These values should suggest an appreciable difference in the crystal structures associated with the four closest local minima to the global minimum, which have been predicted in the search.

4.4.2 MD simulations of the first four structures found in the CSP

We decided to start with an MD simulation of the predicted global minimum structure and of the first three minima, *ak34*, *ak61* and *ai35*, which are the closest in lattice energy to the global minimum, and in the same space group. The MD simulation is performed by using the program DL_MULTI[40], which is an extended version of the program DL_POLY[37], which includes the use of distributed multipoles in the treatment of the electrostatic interactions[41]. The simulations are performed within a rigid molecular model, with a *6-exp* potential having empirically parameters, fitted at room temperature (FIT), which are the same used in the crystal structure prediction of the compound. The simulations are performed in the constant stress ensemble (NPT), using the Berendsen

thermostat, with a relaxation time of 1.0ps and 4.0ps, respectively for the thermostat and the barostat. A cut-off of 12.5 is used for the evaluation of the van der Waals interactions, whereas the electrostatic contribution is treated by a distributed multipole expansion up to hexadecapole order, with an Ewald summation evaluation. The default accuracy (10^{-6}: equation [A.12] in APPENDIX A) is used in the evaluation of the Ewald summation for the multipoles with pole order 1 to 4 in the direct space simulation, in order to avoid problems in the energy conservation, as observed in the energy drift of anisotropic molecules in and out of the cut-off region[35]. For the direct space summation $r_{cut\text{-}off}$ = 12.5Å has been used and for the reciprocal sum, k_{max} corresponds to the larger vector in the reciprocal space. The time step of the simulation is 0.002ps. For structures *ak66-GM* and *ak34* the total simulation time is 800ps, with an equilibration time of 100ps. For the *ak61* and *ai35* structures the total simulation time is 900ps, with a longer equilibration period of 300ps. During the equilibration the velocities of the atoms are rescaled, every 10 ps.

In order to have systems with dimensions as close as possible to each other, the starting points of the MD simulations are the predicted structures in the search, with the following extended cell parameters (table 4.11):

Packing type	***Supercell***	***N° molecules***	***a (Å)***	***b (Å)***	***c (Å)***	***α (°)***	***β (°)***	***γ (°)***
ak66 (GM)	5x5x4	400	38.605	27.505	39.356	90.00	121.05	90.00
ak34	5x5x4	400	34.05	27.510	39.292	90.00	76.31 (103.69)	90.00
ak61	5x6x4	480	33.745	32.952	39.572	90.00	101.10	90.00
ai35	4x7x4	448	36.304	38.318	39.584	90.00	47.01 (132.99)	90.00

Table 4.11: Starting supercell of the MD simulation for the first four local minima predicted in the CSP.

Initially we performed a simulation at three temperatures of 100K, 300K and 400K (close to the melting point) both for the *ak66-GM* structure and for the closest local minimum *ak34*. We are able to reproduce the stability of the interatomic potential in reproducing the experimental structure at different temperatures (up to room temperature) and the effect of the temperature on the lattice parameters, which were calculated in a previous work[26] with an MD simulation of the experimental crystal structure of imidazole. The

latter was performed with the same intermolecular potential (FIT), very similar cut-off and other parameters of the MD simulation, as used in the present study. The starting structures are different: in the previous work, the starting point is the experimental minimised structure of imidazole (with same supercell 5x5x4)[26], whereas in the present study the starting structure is the predicted one, which is very close to the experimental structure but with the molecule in the optimised gas-phase conformation. The results of the simulation at 100K are reported in table 4.12 together with the values obtained in the previous MD study[26].

Structure	***En (kJ mol^{-1})***	***a (Å)***	***b (Å)***	***c (Å)***	***α (°)***	***β (°)***	***γ (°)***	***Vol/Z (Å^3)***
Exp (103K)[6]	-88.1 ± 0.2	7.569	5.366	9.779	90.0	119.1	90.0	86.83
ExMiEx[26] DMAREL (0K)	-78.38	7.72	5.46	9.81	90.0	120.6	90.0	89.00
ExMiEx[26] DL_MULTI (100K)	-76.12	7.783 ±0.002	5.509 ±0.001	9.810 ±0.001	89.998 ±0.020	119.927 ±0.014	90.003 ±0.019	91.13 ±0.14
ExMiOp DMAREL (0K)	-75.14	7.721	5.501	9.839	90.00	121.05	90.00	89.51
ExMiOp DL_MULTI (100K)	-74.52	7.757 ±0.004	5.592 ±0.003	9.824 ±0.003	90.00 ±0.017	120.67 ±0.022	90.00 ±0.019	91.64 ±0.12

Table 4.12: Lattice parameters of the MD simulated *ExMiOp* structure at 100K, compared to the lattice parameters of the MD simulated *ExMiEx* at 100K in a previous work[26] and with the Experimental structure at 103K.

Table 4.12 shows the good agreement between the lattice parameters of the MD simulated structure in the previous work and those from the MD simulated structure in this study. In particular the slight anisotropic thermal expansion along the cell axes ***a*** and ***b*** is well reproduced also in the predicted structure *ak66-GM* (denoted *ExMiOp* in table 4.12).

The change in the lattice parameters with the temperature, for the *ak66-GM* structure and the first closest local minimum *ak34*, are reported in table 4.13.

	En (*kJ mol^{-1}*)	*a* (Å)	*b* (Å)	*c* (Å)	*α* (°)	*β* (°)	*γ* (°)	*Vol/Z* (Å^3)
T = 0K								
ak66-GM	-75.14	7.721	5.501	9.839	90.00	121.05	90.00	89.51
ak34	-74.99	6.810	5.502	9.823	90.00	76.31 ~103.69	90.00	89.96
T = 100K								
ak66-GM	-74.52	7.757	5.592	9.824	90.00	120.67	90.00	91.64
ak34	-74.35	6.881	5.598	9.810	89.99	75.98 (104.02)	90.00	91.66
T = 300K								
ak66-GM	-67.85	7.976	5.701	9.799	89.997	118.25	89.95	98.11
ak34	-67.99	7.928	5.704	9.798	89.99	62.398 (117.60)	90.00	98.13
T = 400K								
ak66-GM	-62.89	8.201	5.718	9.776	90.008	115.12	90.005	103.78
ak34	-62.92	8.198	5.721	9.777	90.00	64.80 (115.20)	89.99	103.73

Table 4.13: Lattice parameters and lattice energy as function of temperature for the MD simulated structures *ak66-GM* and ak34 from the CSP of imidazole.

From the data in table 4.13 we observe continuous stability of the experimental structure, predicted as *ak66-GM*, which undergoes an anisotropic thermal expansion along the ***a*** and ***b*** axes. This behaviour is coherent with the values of the calculated elastic stiffness constant C_{11} and C_{22} (table 4.5), which are much smaller than the value of C_{33}, which represents the major stiffness of the crystal in the direction of the ***c*** axis.

The *ak34* structure, on the contrary is not continuously stable, but undergoes a transformation, at 300K, which makes it relax to the *ak66–GM* structure. This is evident from the strong similarity of the lattice parameters, between the two structures. At 400K

the imidazole structure is over-heated (the melting point of imidazole is ~ 365K) and both *ak66-GM* and *ak34* structures undergo the same thermal expansion. This behaviour confirms the identity of the two structures at room temperature.

From the fact that the closest structure to the global minimum *ak66-GM*, does not relax to the latter one at 100K, we decided to simulate directly at temperature of 300K the structure corresponding to the other two local minima *ak61* and *ai35*. It seems indeed unlikely that the latter can transform at 100K to the experimental structure. The MD simulations for *ak61* and *ai35* are performed for a total time of 900ps with a longer equilibration period of 300ps. The change of the lattice parameters with temperature for the two structures is reported in table 4.14, together with the values of the lattice parameters of the global minimum structure.

	***En* (*kJ mol*$^{-1}$)**	***a* (Å)**	***b* (Å)**	***c* (Å)**	***α* (°)**	***β* (°)**	***γ* (°)**	***Vol/Z* (Å^3)**
T= 0K								
ak66-GM	-75.14	7.721	5.501	9.839	90.00	121.05	90.00	89.51
ak61	-74.66	6.749	5.492	9.893	90.00	101.10	90.00	89.41
ai35	-74.62	9.076	5.474	9.896	90.00	47.01 ~132.99	90.00	89.67
T = 300K								
ak66-GM	-67.85	7.976	5.701	9.799	89.997	118.25	89.95	98.11
ak61	-67.30	7.156	5.699	9.860	89.99	100.73	90.00	98.78
ai35	-61.79	9.915	5.701	9.997	90.00	45.21 {134.79)	90.00	98.76
T = 400K								
ak66-GM	-62.89	8.201	5.718	9.776	90.008	115.12	90.005	103.78
ak61	-67.21	7.678	5.687	9.824	89.99	100.32	89.97	105.50
ai35	-61.66	11.357	5.671	9.824	90.09	41.92 (138.08)	90.03	105.68

Table 4.14: Lattice parameters and lattice energy as function of temperature for the MD simulated structures *ak66-GM* and *ak61* and *ai35* from the CSP of imidazole.

The values of the lattice parameters as function of temperature for the two local minima structures *ak61* and *ai35* (reported in table 4.14) do not relax to the same structure of the

global minimum at high temperatures, as has been observed for *ak34*. The time evolution of the lattice parameters at temperature of 300K for the four structures is reported in figures 4.24-4.28.

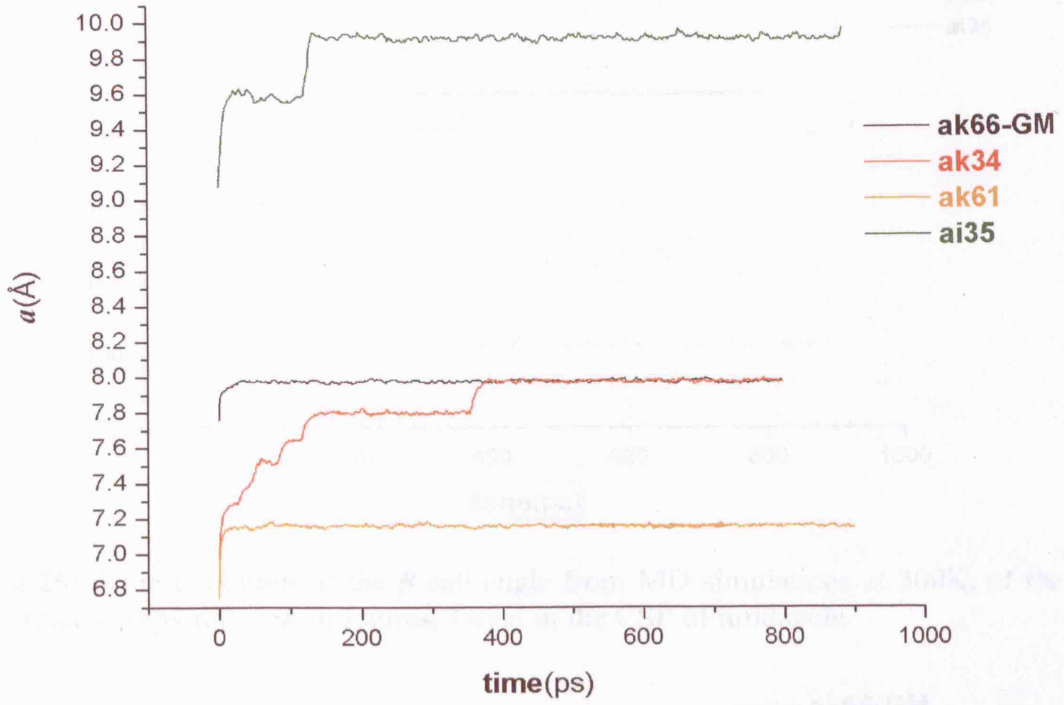


Fig. 4.24: Time evolution of the ***a*** lattice parameter from MD simulations at 300K, of the first four lattice energy minima structures, found in the CSP of imidazole.

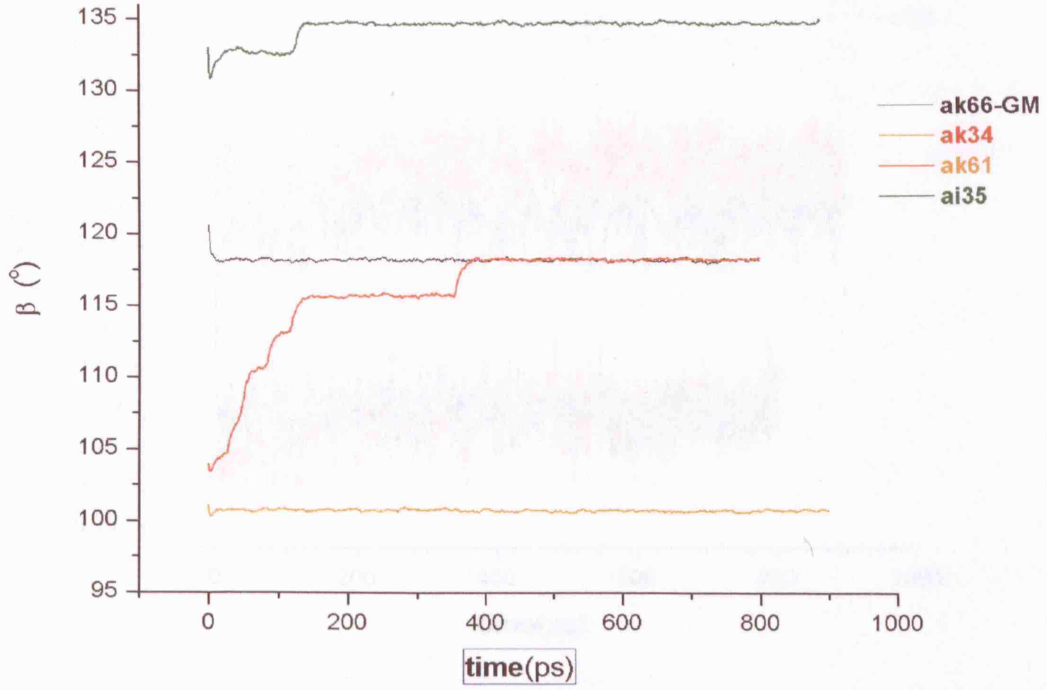


Fig. 4.25: Time evolution of the β cell angle from MD simulations at 300K, of the first four lattice energy minima structures, found in the CSP of imidazole.

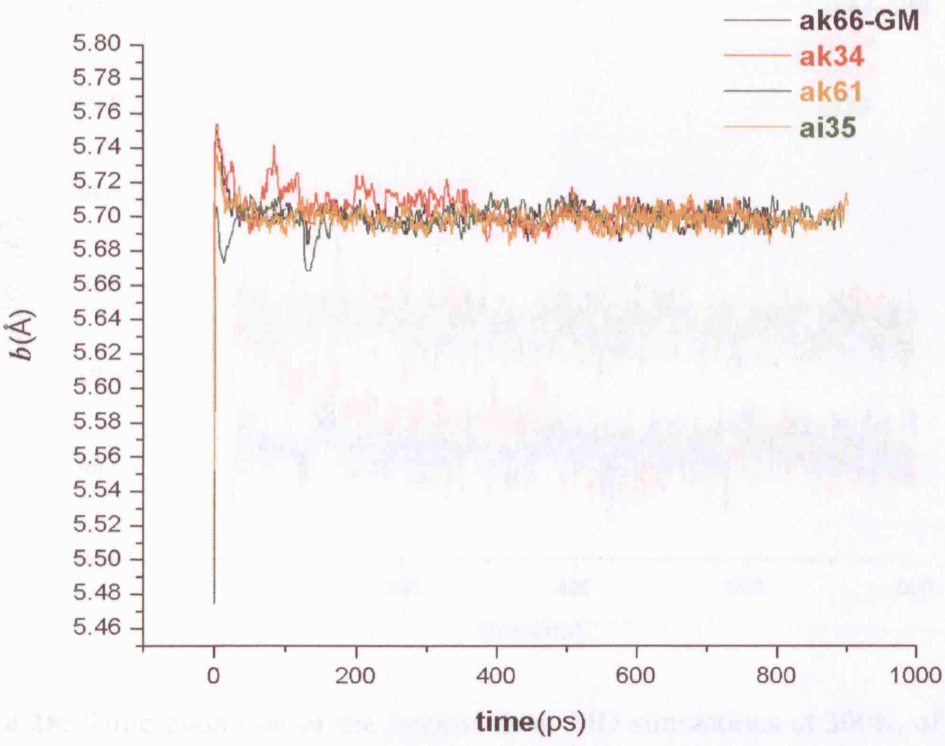


Fig. 4.26: Time evolution of the ***b*** lattice parameter from MD simulations at 300K, of the first four lattice energy minima structures, found in the CSP of imidazole.

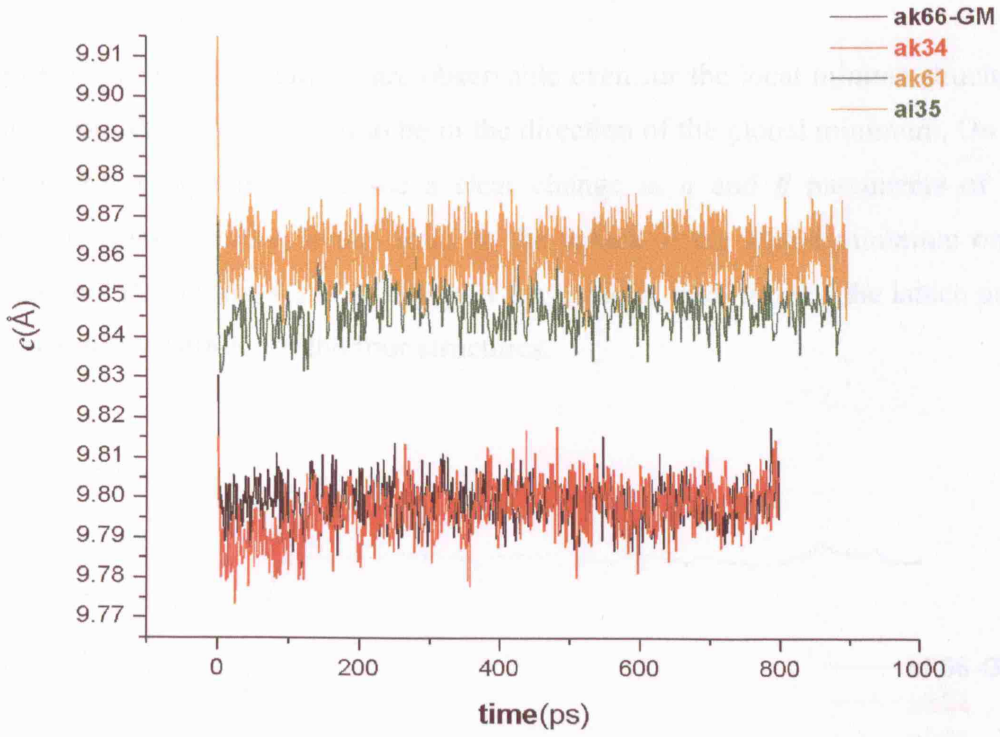


Fig. 4.27: Time evolution of the *c* lattice parameter from MD simulations at 300K, of the first four lattice energy minima structures, found in the CSP of imidazole.

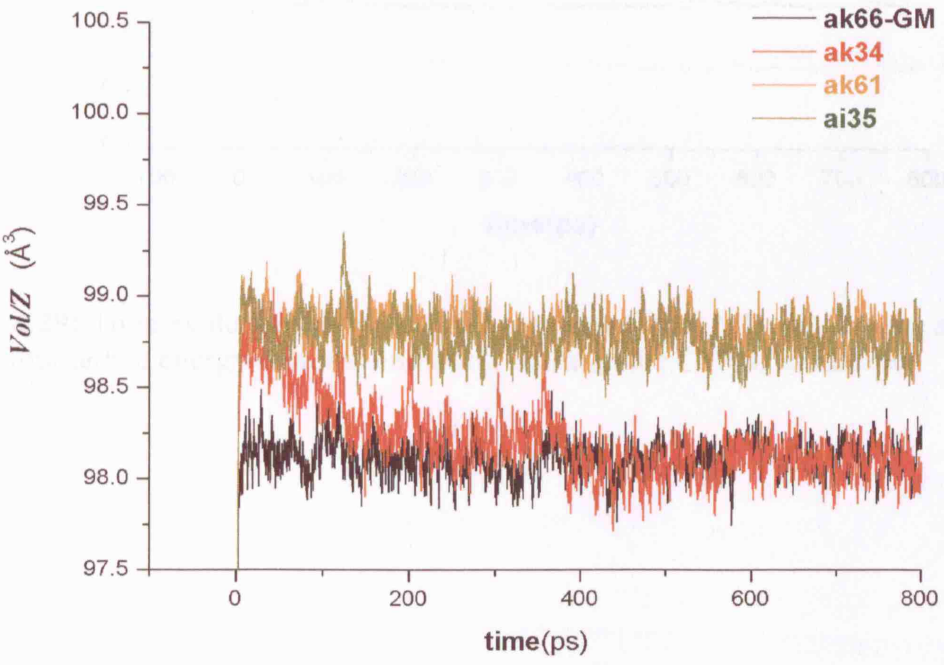


Fig. 4.28: Time evolution of the volume from MD simulations at 300K, of the first four lattice energy minima structures, found in the CSP of imidazole.

At 300K slight transformations are observable even for the local minima structures *ak61* and *ai35*, but they do not seem to be in the direction of the global minimum. On the other hand, at this temperature we see a clear change in ***a*** and ***β*** parameters of the local minimum structure *ak34*, which relax to the values of the global minimum one (figure 4.24 and 4.25). Figures 4.29- 4.33 report the changes with time of the lattice parameters at temperature of 400K for the four structures.

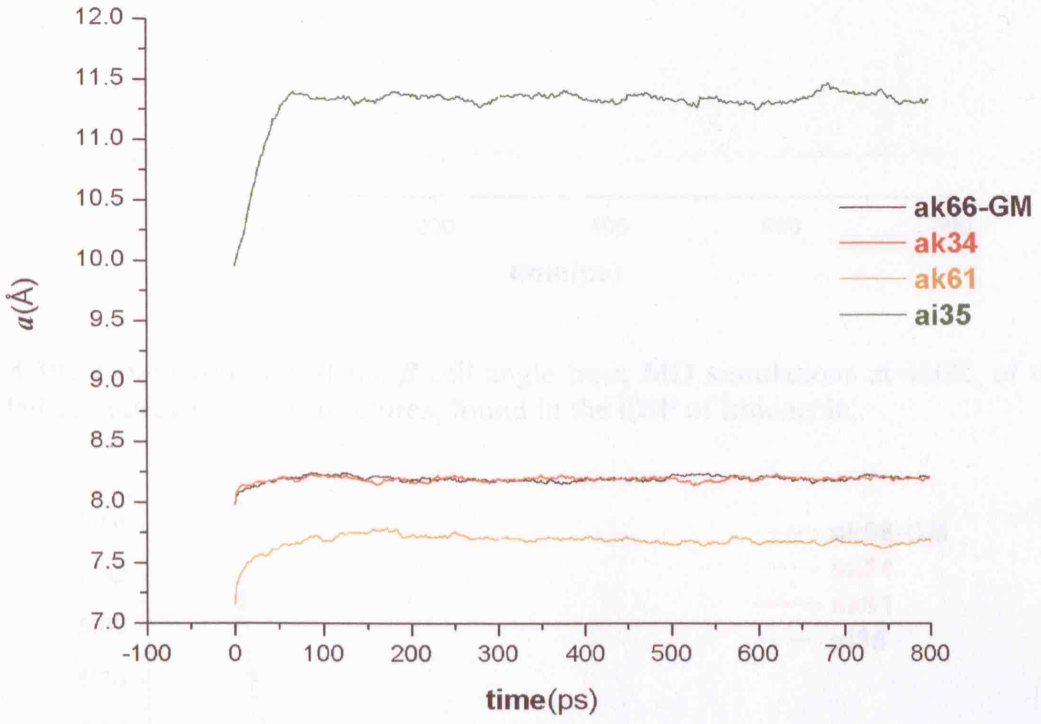


Fig. 4.29: Time evolution of the ***a*** lattice parameter from MD simulations at 400K, of the first four lattice energy minima structures, found in the CSP of imidazole.

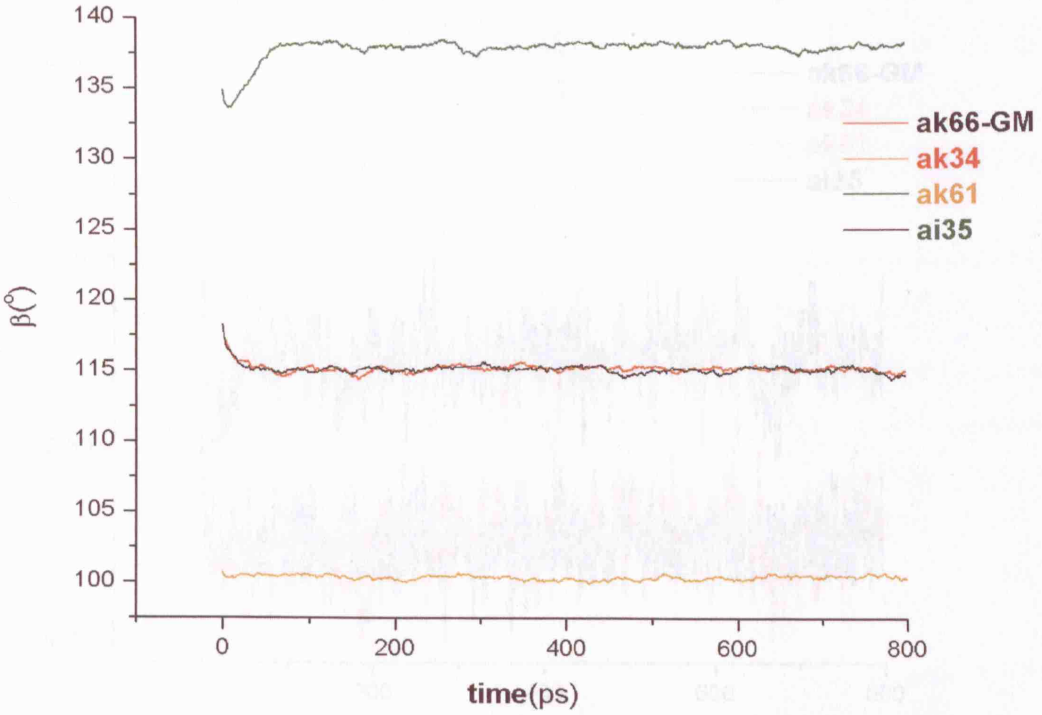


Fig. 4.30: Time evolution of the β cell angle from MD simulations at 400K, of the first four lattice energy minima structures, found in the CSP of imidazole.

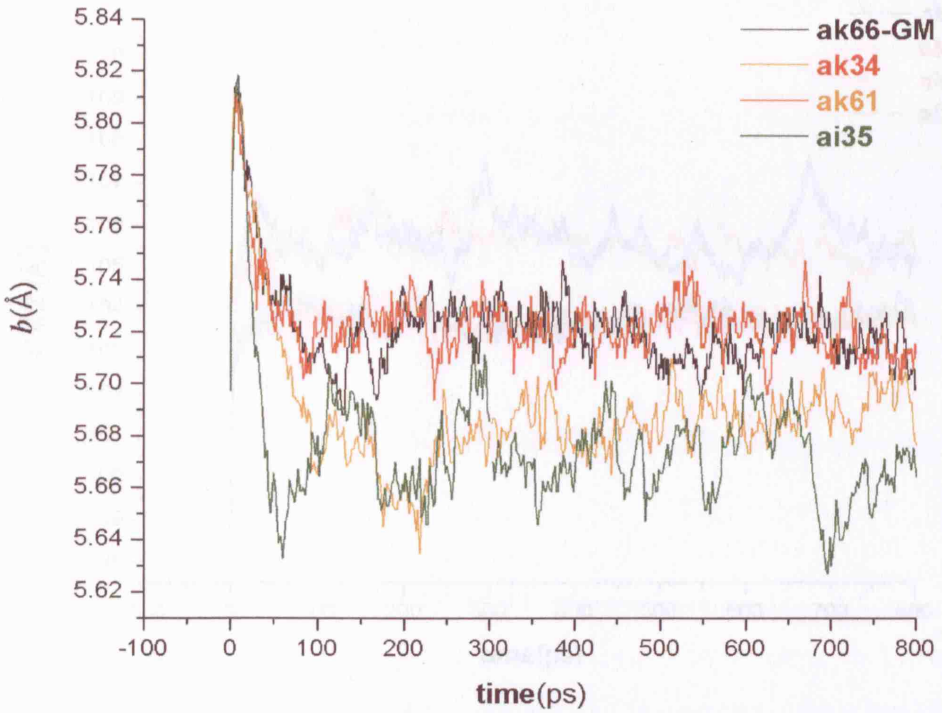


Fig. 4.31: Time evolution of the ***b*** cell parameter from MD simulations at 400K, of the first four lattice energy minima structures, found in the CSP of imidazole.

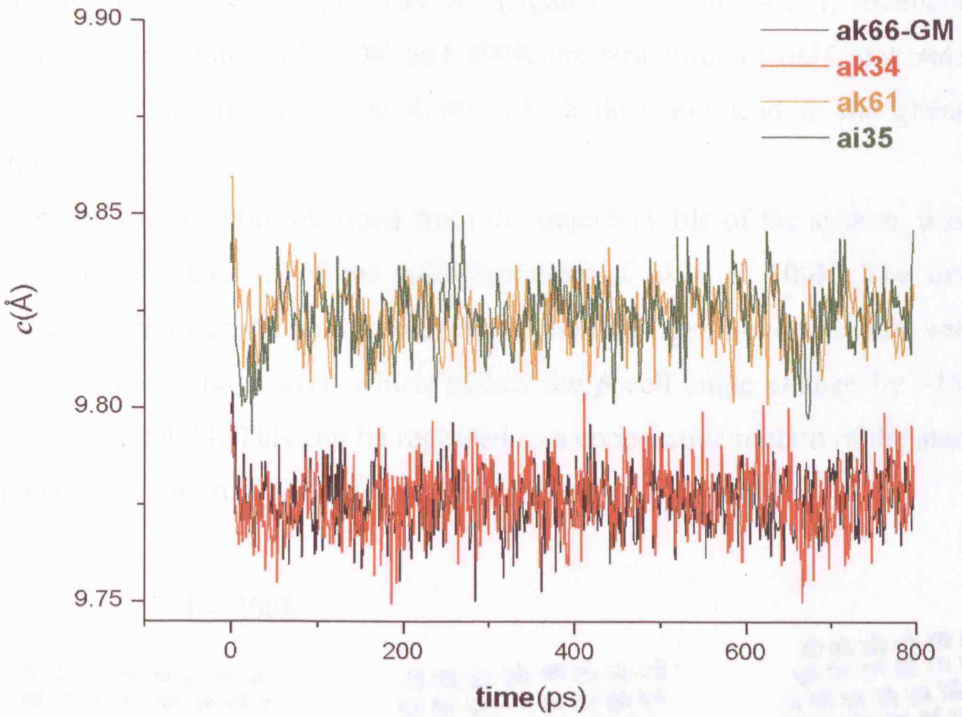


Fig. 4.32: Time evolution of the ***c*** cell parameter from MD simulations at 400K, of the first four lattice energy minima structures, found in the CSP of imidazole.

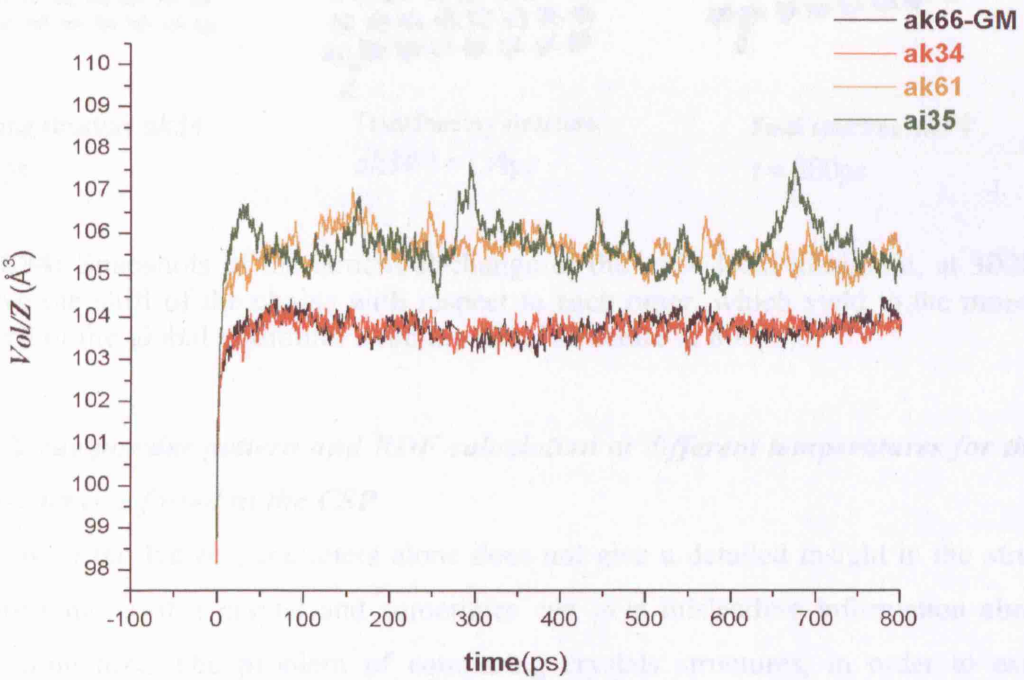


Fig. 4.33: Time evolution of the volume from MD simulations at 400K, of the first four lattice energy minima structures, found in the CSP of imidazole.

While the *ak66-GM* and *ak34*, solely subjected to a thermal expansion, with the volume increasing from ~ 98 $Å^3$ to ~104 $Å^3$ (figures 4.33 and 4.28), maintain eventually identical, at temperature of 300K and 400K the structures of *ai35* and *ak61* undergo a transformation (figures 4.29 and 4.30), which does not lead to the global minimum structure *ak66-GM*.

From the MD animation, obtained from the trajectory file of the system, it is possible to visualise the mechanism of the transformation of *ak34* at 300K. The transformation occurs, through a translational movement of the hydrogen-bonded chains with respect to each other, along the ***a*** axis, which makes the ***β*** cell angle change by ~15 degrees as shown in figure 4.34. This can be regarded as a cooperative motion of the atoms, which is facilitated by the high shear constant along this axis (~ 15 GPa).

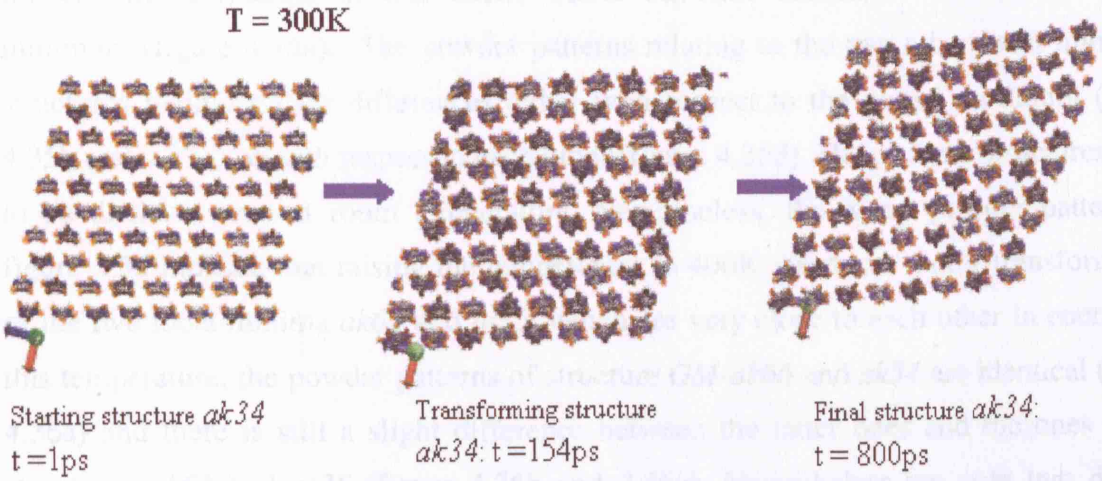


Fig. 4.34: Snapshots of the structural change of the *ak34* local minimum, at 300K. We observe the shift of the chains with respect to each other, which yield to the monoclinic unit cell of the global minimum structure (final structure at 800ps).

4.4.3 X-ray powder pattern and RDF calculation at different temperatures for the first four structures found in the CSP

Analysis of the lattice parameters alone does not give a detailed insight in the structural transformations of a crystal and sometimes can give misleading information about the phase transition. The problem of comparing crystals structures, in order to establish similarities and differences is widely debated and it becomes even harder, when using MD simulated systems with large simulation cells, which contain intrinsic molecular dynamical motions. One encouraging and useful method consists of calculating X-ray powder patterns from the molecular dynamic, by using the atomic positions in the

trajectory of the system[42-44]. In order to avoid the fluctuations of the lattice parameters included in the trajectory of the system in the NPT ensemble MD simulations, we performed for each structure, at each temperature a short simulation in NVE ensemble for 40 ps after an equilibration of 200ps with a time step of 0.002 ps. The starting point of the simulation in NVE ensemble is the configuration of the system having the cell parameters as close as possible to the average structure, obtained from the NPT ensemble simulation. The method and the program applied to calculate the X-ray powder patterns are described in chapter 3 section 3.6.

Figures 4.35 and 4.36 show the calculated powder patterns for the four structures at both 300K and 400K.

At 300K the X-ray powder patterns of the global minimum structure *GM-ak66* and of the close local minimum *ak34* match perfectly, giving further evidence for the thermally induced transformation of this latter, whose structure relaxes to that of the global minimum (figure 4.35a). The powder patterns relating to the two other local minimum structures maintain their differences either with respect to the global minimum (figure 4.35b and 4.35c) or with respect to each other (figure 4.35d). Hence these structures seem to be distinct, even at room temperature. Nevertheless, the X-ray powder patterns in figure 4.36 indicate that raising the temperature to 400K, produces a new transformation of the two local minima *ak61* and *ai35*, which are very close to each other in energy. At this temperature, the powder patterns of structure *GM-ak66* and *ak34* are identical (figure 4.36a) and there is still a slight difference between the latter ones and the ones of the structures *ak61* and *ai35* (figure 4.36b and 4.36c). Nevertheless we note two distinct points: first, the similarity between all the powder patterns has increased compared to those at 300K and secondly, the powder patterns of the two local minima *ak61* and *ai35* are identical at 400K, indicating that these structures become effectively the same at this temperature (figure 4.36d). Therefore is clear from the analysis of the calculated X-ray powder pattern that thermal effect at temperatures $T \geq 300K$ results in a process of "coalescence" of the local minima. Four distinct minima at $T < 300K$ are resolved in sole two minima at 400K, with *GM-ak66* and *ak34* representing the global minimum and *ak61* and *ai35* the closest local minimum. The ranking of the minima has not been modified with temperature, with *GM-ak66* - *ak34* still corresponding to the global minimum in the internal and total energy and *ak61- ai35* corresponding to the closest local minimum (figure 4.37).

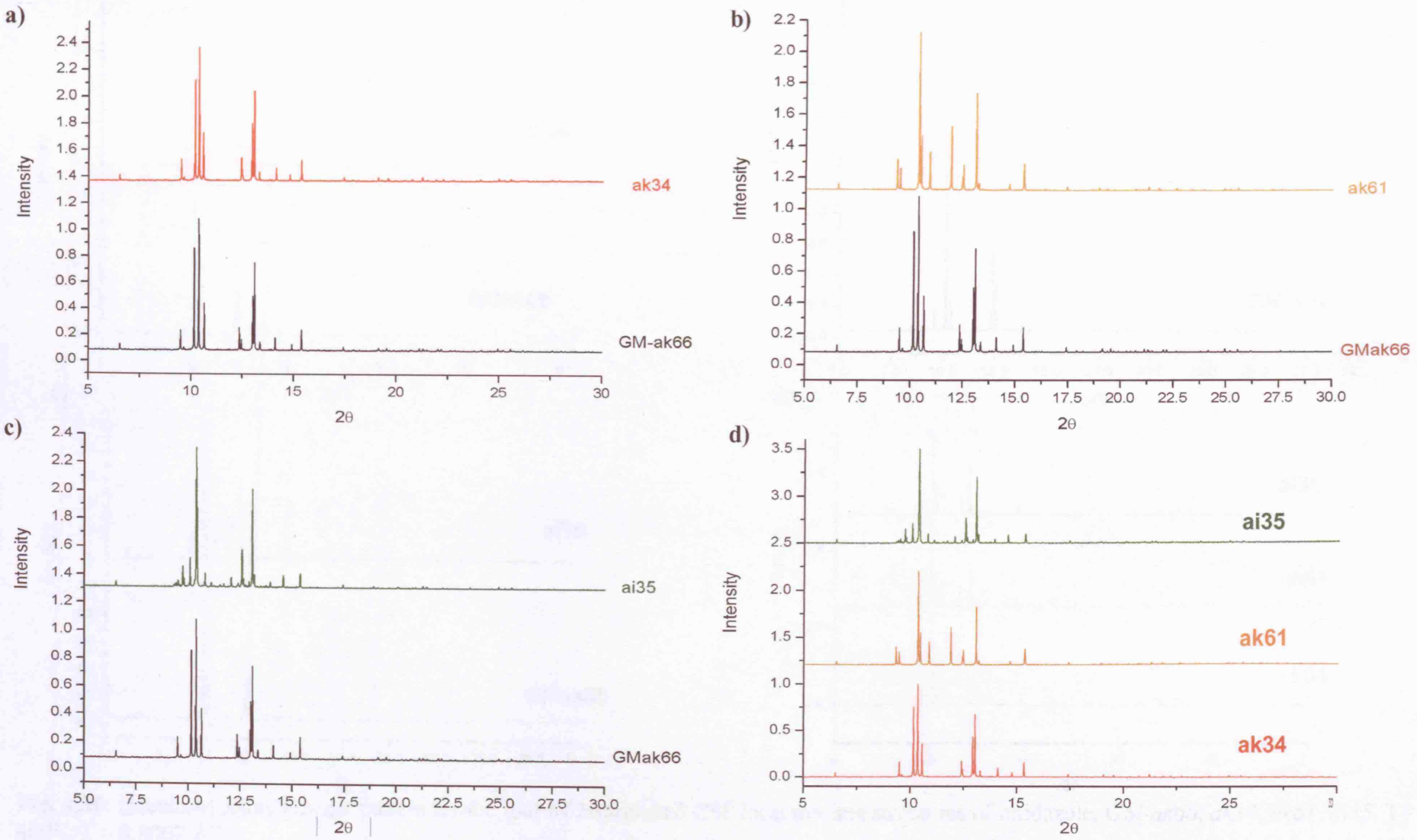


Fig. 4.35: calculated X-ray powder pattern for the four MD simulated CSP local minima structures of imidazole, *GM-ak66*, *ak34*, *ak61*, *ai35*. T= 300K. $\lambda = 0.8002$ Å^{-1}.

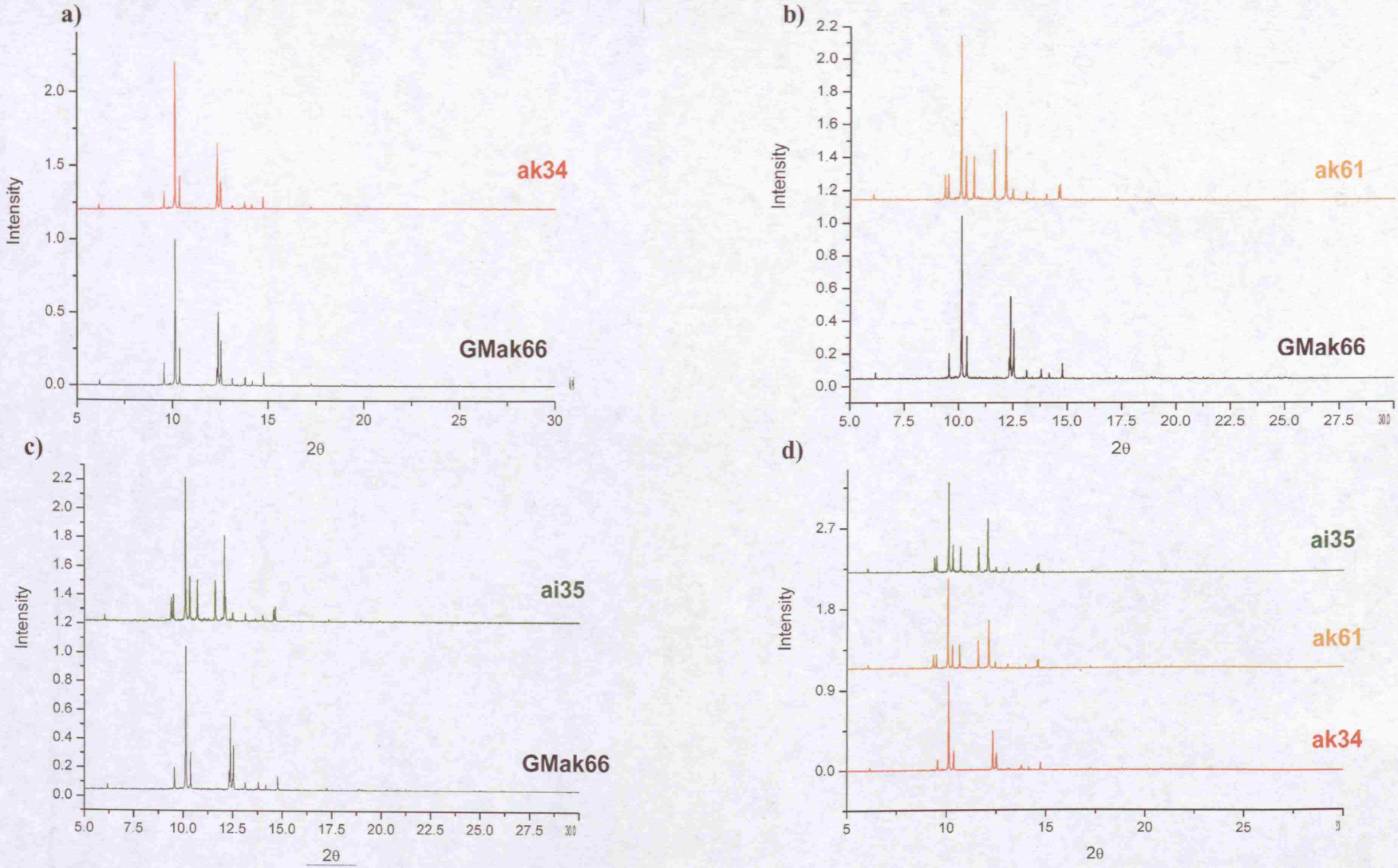


Fig. 4.36: calculated X-ray powder pattern for the four MD simulated CSP local minima structures of imidazole, *GM-ak66*, *ak34*, *ak61*, *ai35*. T= 400K. $\lambda = 0.8002$ Å^{-1}.

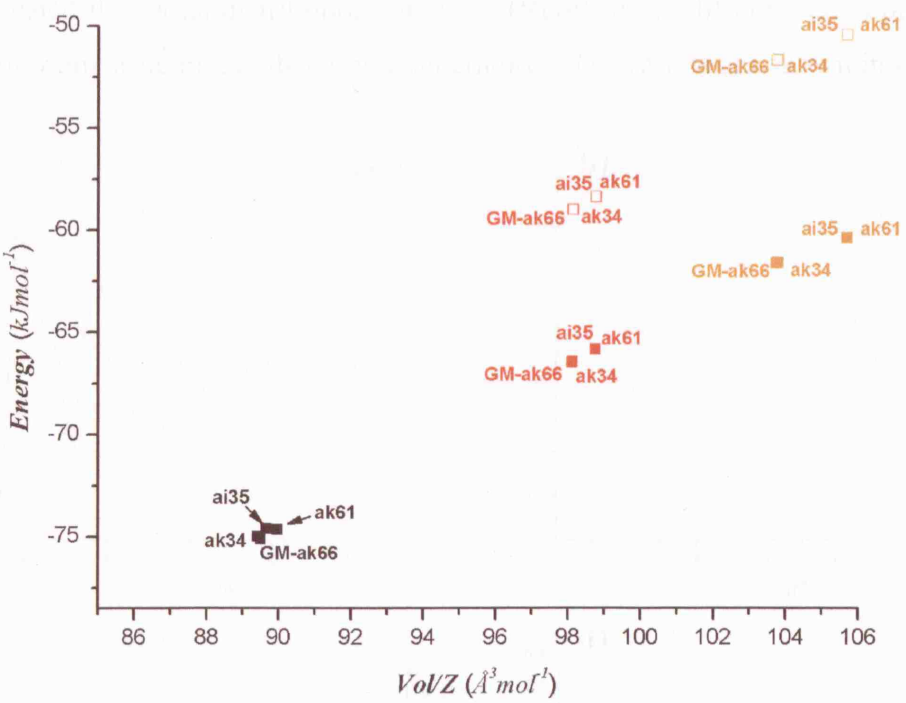


Fig. 4.37: Lattice energy (solid squares) and total energy (empty squares) of the predicted global minimum (*GM-ak66*) and the closest three predicted local minima (*ak61*, *ak34*, *ai35*) at temperature of 0K (black square), 300K (red square), and 400K (orange square).

In order to gain a better understanding of the structural differences of the local minima, we calculated the radial distribution functions (RDF) of the different atom pairs, for each local minimum structure, at different temperatures. The results are shown in figure 4.39.

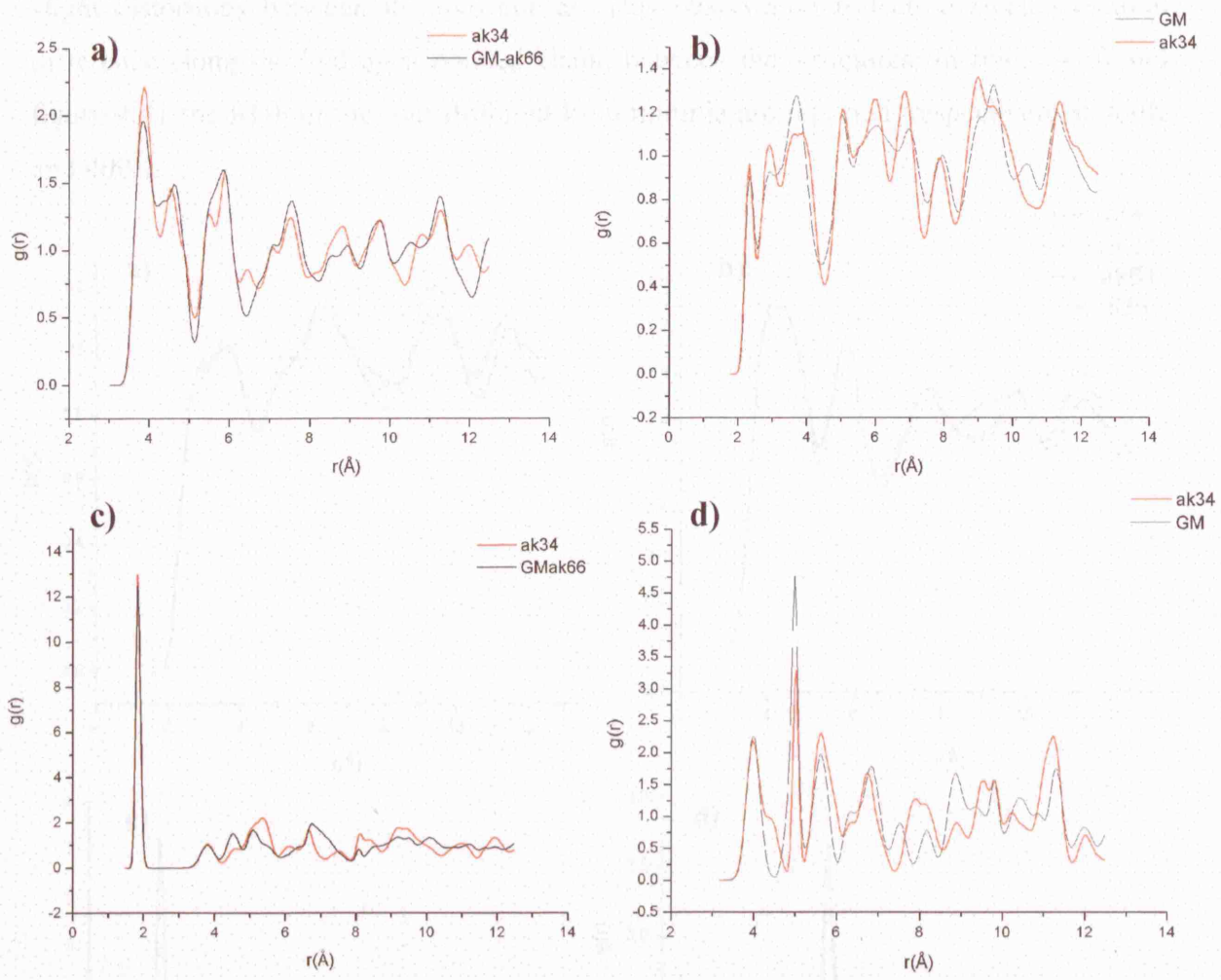


Fig. 4.38: Radial distribution function (RDF) plots of the global minimum structure (*GMak66*: black curve) and the closest local minimum in lattice energy (*ak34*: orange curve), at 100K. **a)** H1-H1 RDF, **b)** C-C, **c)** N2-H2, **d)** N2-N2.

The structural differences between the two crystal structures *GMak66* and *ak34* at low temperature are reflected in the pair distribution function RDF of the different atomic species in the molecule. We considered in particular the H1-H1, C1-C1, N2-N2 and N2-H2 correlations, which mark the differences in the relative orientations and positions of the molecules in the structure. As we can see in figure 4.39 the relative positions of the atoms are relatively similar for the first neighbours (up to 4 Å) and differ for bigger distances. The difference in the profiles of the RDF is still small, especially for the N2-H2 pairs (figure 4.39c), which lie along the hydrogen bond chain. The biggest differences are observable in the relative positions of C1 (figure 4.39b) atoms with respect to each

other, in the intermolecular interaction and in those of the H1 atoms (figure 4.39a). An apparent big difference is also observable in the RDF of N2-N2 pairs, but a careful examination reveals very similar profiles of the RDF curves, even at large distances, with slight distortions between the two curves. This observation reflects a small structural difference along the hydrogen bonded chain, between the structures. In figure 4.40 and figure 4.41 the RDF of the four different local minima are reported, respectively at 300K and 400K.

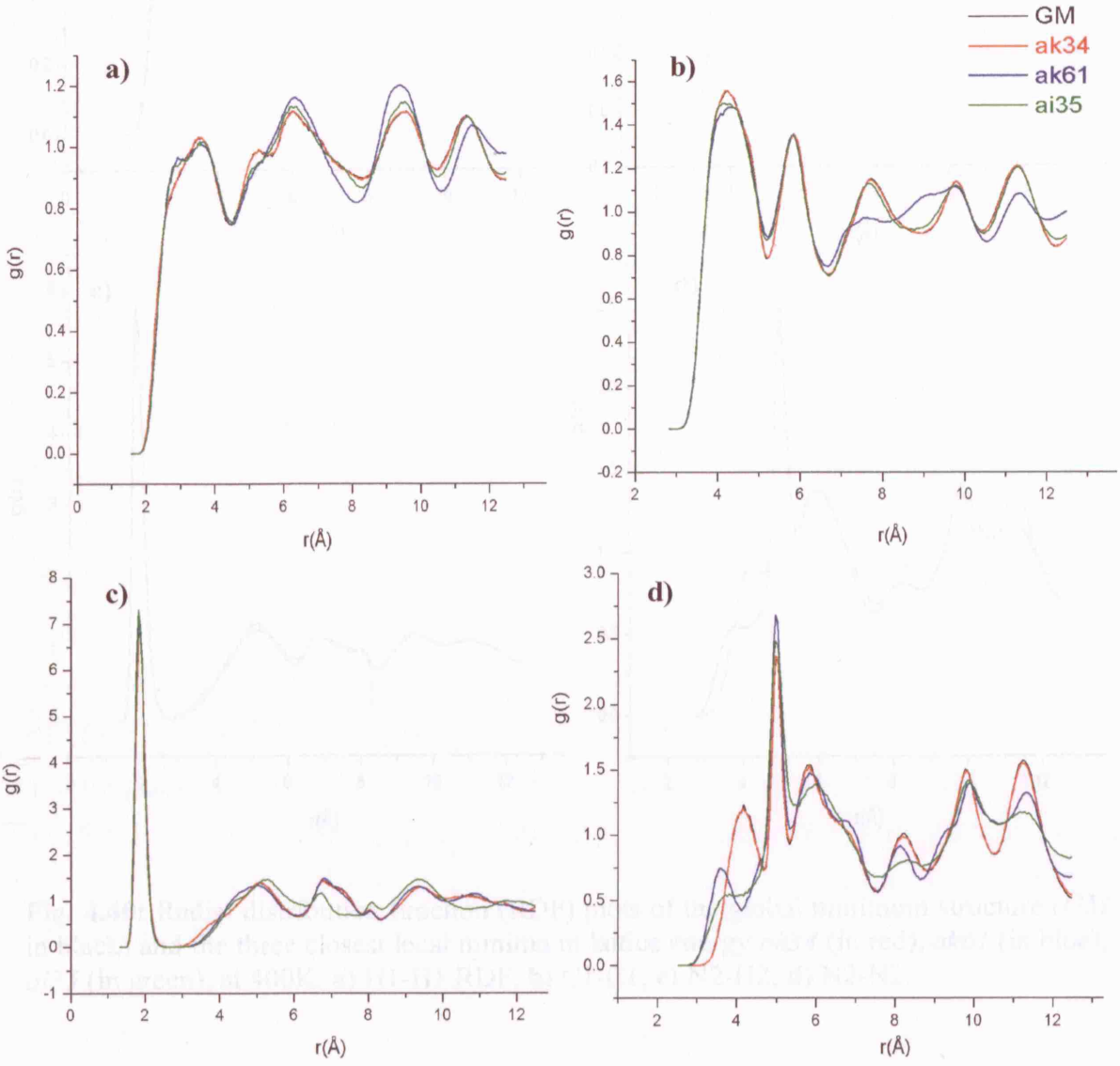


Fig. 4.39: Radial distribution function (RDF) plots of the global minimum structure (*GM* in black) and the three closest local minima in lattice energy *ak34* (in red), *ak61* (in blue), *ai35* (in green), at 300K. **a)** H1-H1 RDF, **b)** C1-C1, **c)** N2-H2, **d)** N2-N2.

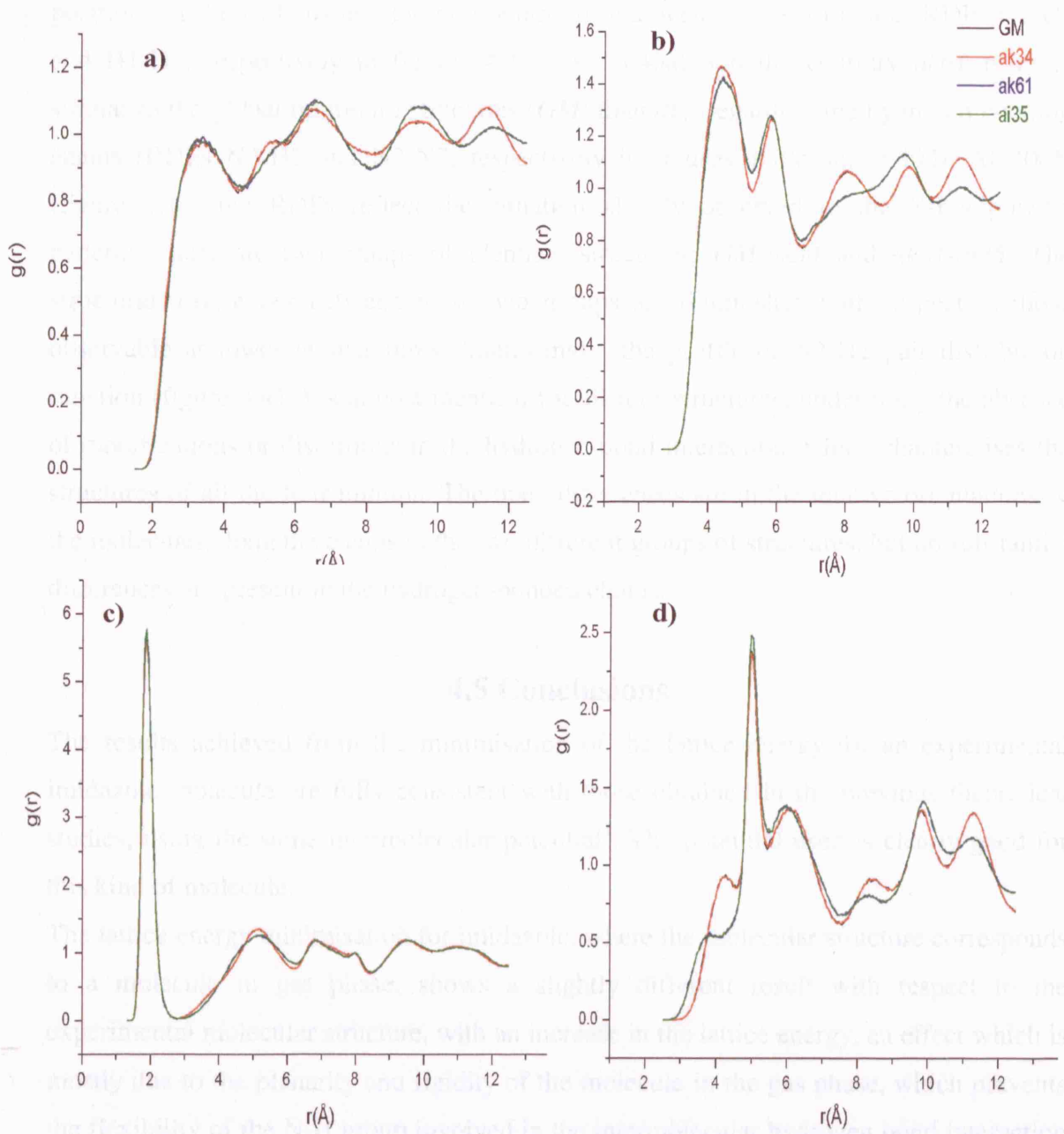


Fig. 4.40: Radial distribution function (RDF) plots of the global minimum structure (*GM* in black) and the three closest local minima in lattice energy *ak34* (in red), *ak61* (in blue), *ai35* (in green), at 400K. **a)** H1-H1 RDF, **b)** C1-C1, **c)** N2-H2, **d)** N2-N2.

At 300K, we observe that the RDF of *GM* and *ak34* is the same for all of the pairs of atoms in imidazole molecule. On the contrary the RDF profiles of the structure *ak61* and *ai35* differ from the former two and also differ from each other. This behaviour is in good agreement with the results obtained from the calculation of the X-ray powder patterns of the four structures. Interesting RDF profiles reveal the structure *ai35* to be more similar to the global minimum structures (*GM*) than the structure *ak61*, regarding the relative

positions of the carbons and the non-polarised hydrogens in the structure (RDFs C1-C1 and H1-H1, respectively in figures 4.40b and 4.40a). On the contrary *ak61* is more similar to the global minimum structures (*GM*) than *ai3,* regarding the hydrogen bonding chains (RDFs N2-H2 and N2-N2, respectively in figures 4.40c and 4.40d). At 400K (figure 4.41) the RDFs reflect the situation already observed in the X-ray powder patterns: there are two groups of identical structures: *GM-ak34* and *ak61-ai35*. The structural differences between these two groups are diminished with respect to those observable at lower temperatures. Interestingly, the profile of N2-H2 pair distribution function (figure 4.41c) is almost identical for all four structures, underlining the absence of modifications or distortions in the hydrogen bond interaction, which characterises the structures of all the four minima. The main differences are in the relative orientations of the molecules along the chains in the two different groups of structures, but no substantial differences are present in the hydrogen-bonded chains.

4.5 Conclusions

The results achieved from the minimisation of the lattice energy for an experimental imidazole molecule are fully consistent with those obtained in the previous theoretical studies, using the same intermolecular potential[5]. The potential used is clearly good for this kind of molecule.

The lattice energy minimisation for imidazole, where the molecular structure corresponds to a molecule in gas phase, shows a slightly different result with respect to the experimental molecular structure, with an increase in the lattice energy, an effect which is mostly due to the planarity and rigidity of the molecule in the gas phase, which prevents the flexibility of the N-H group involved in the intermolecular hydrogen bond interaction inside the lattice. Nevertheless the tiny differences between the experimental and the "gas phase" *ab initio* optimised geometry, have been demonstrated not to dramatically affect the reliability of the model for such a kind of rigid molecule.

The differences between the values of the lattice energy in both cases with respect to the sublimation enthalpy might be partially due to the potential used, but more likely are caused by the presence of an distortions in the structure produced by the thermal effect[45].

Temperature effects also determine the difference between the measured and calculated phonons frequencies, which in the case of the experimental molecular structure crystal lattice reproduce exactly the values obtained in the previous calculations[5].

The estimate of the elastic constants, either for the optimised molecular structure or for the experimental one, shows an anisotropic behaviour of the crystal under compression, which is stronger along the *c* axis of the unit cell.

The calculation of the lattice energy, using only the atomic point charges (CHelpG method) gives results for the energy and the **c** value of the lattice parameters, which are very far from the values obtained experimentally. This model does not reproduce the experimental crystal structure satisfactorily. The elastic constants have lower values with respect to the values calculated with the DMA method (hence the crystal has a lower mechanical stability) and also the estimate of the phonon frequencies is not accurate (RMS residual is the largest among the different methods used). The principal reason is the presence of the hydrogen bond between the molecules, and consequently the important contribution of the electrostatic interactions, which is modelled better using the multipole expansion of the atomic charges.

The crystal structure prediction using intermolecular potentials with the FIT parameters and DMA of the electronic charges on the atoms was successful. The experimental structure was predicted as the global minimum in the lattice energy and the total free energy with very good estimates of the lattice parameters and HB lengths.

The prediction of likely polymorphic structures is more difficult, because the presence of several structures in a small range of lattice energy. Some of these structures (in the same space group of the experimental structure and in a range of lattice energy ~ k_BT) are identical to the experimental structure if we take into account thermal effects and phonon contributions in a more realistic approximation. Three other structures, with different symmetry, are good candidates as polymorphs of imidazole, considering their thermodynamic properties and volume growth rates. An analysis of the hydrogen bond motif of these structures highlights the presence of the same type of hydrogen bond motif in all of the structures. This implies a low potential energy barrier for the transition from a structure at higher energy to the global minimum structure, which makes the transition from a metastable to the stable form relatively facile. Hence the presence of possible polymorphs for this compound seems to be unlikely.

It was generally observed that there was an increase in the lattice energy for structures with low P.C. and short HB length (< 1.83Å) and for structures with high P.C. values, but also long HB length (> 1.92 Å). In this latter case, the reason for the increase in the energy could be the presence of a extensive stress in the structure forcing it to have dense packing, despite the presence of a long HB distance.

An attempt to include thermal effects was considered for the first three local minima in lattice energy, by performing a MD study of the four structures (the global minimum and the other three local minima) at high temperatures (300K and 400K). The MD simulations are possible because of the small differences in the free energies of the structures (table 4.7), which make us suppose relatively low energy barriers involved in the structural differences of these local minima. These low energy-barrier phase transitions do not involve breaking and reconstructing of intra/inter-molecular bonds, typical of reconstructive phase transitions and solid state reactions like proton transfer[46],as in the case of terephthalic compound[47]. Indeed, some structures, such as *av6* (figure 4.12), are present among all the local minima of imidazole within 5kJ mol^{-1} from the global minimum, in which, unlike the global and the other local minima previously considered in the MD study, the adjacent chains have iso-oriented dipole moments, given by the sequence of the N-H polar groups, pointing all in one direction. This structural property makes this crystal polar (as schematically illustrated in figure 4.42a). A transition from a polar to a non-polar structure (like that of the global minimum) involves the breaking of strong hydrogen bond interactions, and since it is a very energetic phenomenon, which cannot be simulated through an MD simulation.

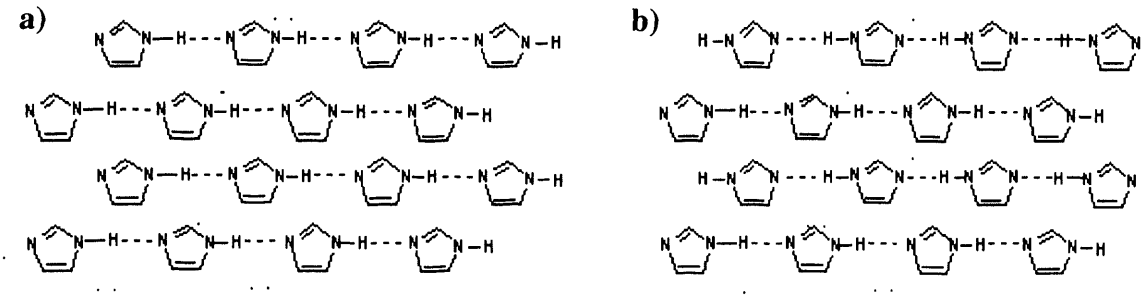


Fig. 4.41: Schematic illustration of the distinct polar structure **(a)** and *non*-polar structures **(b)** for imidazole compound. The polarity is given by the hydrogen bond interaction **N-H --- N** along the chains.

Other more recent computational techniques such as Metadynamics[48,49], are able to simulate this kind of transformation, in particular for the imidazole compounds[50], by giving detailed information about the mechanism of transformation.

As we could expect, from the MD simulations of the four different local minima structures we see that thermal effects play an important role, which allows the system to overcome the relatively small energy barriers, leading to the coalescence of local minima,

which result as different structures at low temperature. The analysis of the X-ray powder patterns has proved to be more powerful and efficient than the sole analysis of the lattice parameter change in indicating structural transformations among local minima. Finally, detailed insights in the structure of the different local minima can be given by the analysis of the pair radial distribution functions (RDF), which can show local similarities and differences among the structures of the different local minima. This study confirms the possibility of using MD simulation in attempting to solve the problem of the large variety of local minima in a computational crystal structure prediction, which does not include temperature effects, as already suggested in a previous work[36]. The calculation of X-ray powder patterns and of the RDF of the MD structures can help in detecting structural similarities and differences among the different local minima obtained in the computational search.

Overall this work represented a very good success in the prediction of crystal structure of simple rigid organic molecules, and has shown all the possible useful information, which can obtain by the synergy of several computational techniques, under the assumption of the use of a reliable and accurate intermolecular potential.

4.6 References

(1) Bernstein, J. *Polymorphism in Molecular Crystals*, 2002.

(2) Gavezzotti, A. *Acc.Chem. Res.* **1994**, *27*, 309.

(3) Dunitz, J. D. *Chem. Comm.* **2003**, 545.

(4) Allen, F. H.; Motherwell, W. D. S. *Acta Cryst. Sec. B* **2002**, *58*, 407.

(5) Day, G. M.; Price, S. L.; Leslie, M. *J. Phys. Chem. B* **2003**, *107*, 10919.

(6) McMullan, R. K. E., J. ; Ruble, J. R. ; Craven, B. M. *Acta Cryst. Sec. B* **1979**, *35*, 688.

(7) Moeller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(8) Frisch, M. J. T., G.W.; Schlegel,H.N.; Scuderia,G.E.; Robb,M.A.; Chesseman,J.R.; Zakrzewski,V.G.; Montgomery,J.A.; Stratmann,R.E.; Burant,J.C.; Dapprich,S.; Millam,J.M.; Daniels,A.D.; Kudin,K.N.; Strain,M.C.; Farkas,O.; Tommasi,J.; Barone,V.; Cossi,M.; Cammi,R.; Mennucci,B.; Pomelli,C.; Adamo,C.; Clifford, S.; Ochterski,J.; Petersson,G.A.; Ayalla,P.Y.; Cui,Q.; Morokuma,K.; Malick,D.K.; Rabuck,A.D.; Raghavachari,K; Foresman,J.B.; Cioslowski,J.; Ortiz,J.V.; Stefanov,B.B.; Liu,G.; Liashenko,A.; Piskorz,P.; Komaromi,L.; Gomperts,R.; Martin,R.L.; Fox,D.J.; Keith,T.; Al-Laham,M.A.; Peng,C.Y.; Nanayakkara,A.; Gonzalez,C.; Challacombe,M.; Gill,P.M.W.; Johnson,B.G.; Chen,W.; Wong,M.W.; Andres,J.L.; Head-Gordon,M.; Replogle,E.S.; Pople,J.A.; *GAUSSIAN 98*; Gaussian Inc.: Pittsburgh, 1998.

(9) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications: New York, 1996.

(10) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

(11) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(12) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(13) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2004**, *4*, 1327.

(14) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628.

(15) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.

(16) Holden, J. R.; Du, Z. Y.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.

(17) Spek, A. L. *J. Appl. Crystallogr.* **2003**, *36*, 7.

(18) Epstein, J.; Ruble, J. R.; Craven, B. M. *Acta Cryst. Sec. B* **1982**, *38*, 140.

(19) Munch, W.; Kreuer, K. D.; Silvestri, W.; Maier, J.; Seifert, G. *Solid State Ion.* **2001**, *145*, 437.

(20) Accelrys, I. Cerius2; 4.0 ed.; Accelrys Inc: San Diego, 2000.

(21) Jimenez, P. R., M.V. ; Turrion, C. *J.Chem. Thermodyn.* **1987**, *19*, 985.

(22) Gavezzotti, A. *Cryst. Rev.* **1998**, *7*, 5.

(23) Gavezzotti, A. *Theoretical Aspects and Computer modeling of the Molecular Solid State*; John Wiley & Sons: Milan, 1997.

(24) Majoube, M.; Vergoten, G. *J. Chem. Phys.* **1982**, *76*, 2838.

(25) Perchard, C.; Novak, A. *J. Chem. Phys.* **1968**, *48*, 3079.

(26) Gray, A. E.; Day, G. M.; Leslie, M.; Price, S. L. *Mol. Phys.* **2004**, *102*, 1067.

(27) Cammers, A.; Parkin, S. *CrystEngComm* **2004**, *6(29)*, 168.

(28) Bruno, J.; Cole, J. C.; Edgington, P. R.; Kessler, M. K.; Macrae, C. F.; McCabe, P.; J., P.; Taylor, R. *Acta Cryst. Sec. B* **2002**, *58*, 389.

(29) Kitaigorodsky, A. I. *Molecular Crystals and Molecules*New York, 1973.

(30) Gale, J. D.; Rohl, A. *Mol. Simul.* **2003**, *29*, 291.

(31) Fleming, S. GDIS; 0.80 ed.; Curtin University of Technology: Perth, 2003.
(32) Price, S. L. *Adv. Drug Deliv. Rev.* **2004**, *56*, 301.
(33) Dove, M. T. *Introduction to lattice dynamics*; Cambridge University Press, 1993.
(34) Gavezzotti, A.; Filippini, G.; Kroon, J.; vanEijck, B. P.; Klewinghaus, P. *Chem.-Eur. J.* **1997**, *3*, 893.
(35) Gray, A. E., Day, G. M. , Leslie, M. , Price, S. L.
(36) Gavezzotti, A. *J.Am.Chem. Soc.* **2000**, *122*, 10724.
(37) Smith, W.; Forester, T. R. *J.Mol.Graphics* **1996**, *14*, 136.
(38) Davey, R. J.; Maginn, S. J.; Andrews, S. J.; Black, S. N.; Buckley, A. M.; Cottier, D.; Dempsey, P.; Plowman, R.; Rout, J. E.; Stanley, D. R.; Taylor, A. *J. Chem. Soc.-Faraday Trans.* **1994**, *90*, 1003.
(39) Chisholm, J. A.; Motherwell, S. W. D. *J. Appl. Crystallogr.* **2005**, *38*, 228.
(40) Leslie, M. *Mol. Phys.* **2004**, *(in Press)*.
(41) Stone, A. J.; Alderton, M. *Mol.Phys.* **1985**, *56*, 1047.
(42) Kihara, K.; Matsui, N. *Phys. Chem. Miner.* **2003**, *30*, 345.
(43) Gavezzotti, A. *Molecular Aggregation*; Oxford Science Publications, 2007.
(44) McGreevy, R. L. *J. Phys.-Condes. Matter* **2001**, *13*, R877.
(45) Beyer, T.; Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2001**, *123*, 5086.
(46) Dunitz, J. D. *Pure & Appl. Chem.* **1991**, *63*, 177.
(47) Fischer, P.; Zolliker, P.; Meier, B. H.; Ernst, R. R.; Hewat, A. W.; Jorgensen, J. D.; Rotella, F. J. *J. Solid State Chem.* **1986**, *61*, 109.
(48) Buch, V.; Martonak, R.; Parrinello, M. *J. Chem. Phys.* **2005**, *123*.
(49) Martonak, R.; Laio, A.; Bernasconi, M.; Ceriani, C.; Raiteri, P.; Zipoli, F.; Parrinello, M. *Z. Kristall.* **2005**, *220*, 489.
(50) Iannuzzi, M.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *93*.

Chapter 5

AZETIDINE CRYSTAL STRUCTURE PREDICTION: Blind test 2004

5.1 Introduction

5.1.1 The Cambridge Crystallographic Structure Database Blind tests

An important contribution in testing the progress reached in the field of crystal structure prediction (CSP) is given by the Cambridge Crystallographic Structures Database (CSD)[1], which has been organising a series of competing "blind tests", which are challenges to predict the crystal structures of compounds, whose structures are unpublished, starting solely from the chemical formula of the compound.

So far three blind tests had been made, in 1999[2], 2001[3] and 2004[4], to which participants from all over the world have taken part, in the attempt to predict crystal structure of organic molecules. A group of usually three candidate molecules is chosen with the following rules, to give a reasonable chance of success within the practical limits of the computer programs:

- Maximum number of atoms, including H, not to exceed 40.
- Space groups to be one of the most common in the organic solid state: $P2_1/c$, *P-1*, $P2_1$, $P2_12_12_1$, *C2/c*, *Cc*, *Pbca*, *Pbcn*, $Pna2_1$, *C2*.
- One molecule in the asymmetric unit: Z`=1
- No solvent molecules in the crystal structure
- No co-crystals
- No disorder in the structures
- Hydrogen bonds located experimentally in the determined structure.

Usually the three molecules chosen in the first two blind tests belonged to three different categories[4]:

a) small rigid molecules with a maximum number of 25 atoms, containing solely species like: C, O, N, H.

b) rigid molecules containing functional groups, that present a challenge for the method, but with less than 40 atoms

c) molecules with torsional degrees of freedom.

In the last blind test the restriction to crystals with only Z`=1 was extended to the condition of Z`<3, given the discrete rate of success observed in the previous two blind tests[4], in predicting structures with Z`=1. Each participant can submit three choices as expected crystal structure, for each molecule.

Several search methods, listed in section 1.6, were used by the participants in the three blind tests. All these methods consist of three main steps: building of all the possible local minima, energy minimisation of the structures and clustering to eliminate redundant equivalent structures, plus additional criteria to allow a choice of the three low energy structures to be submitted. The methods also can be divided into two types of modelling procedure. The first[5] lets the molecule change its internal conformation, considering the internal molecular degrees of freedom as variables in the minimisation procedure, while the other uses a rigid body approximation. For the latter, in the case of flexible molecules, a separate search for each low energy conformer is performed after a careful conformational analysis of the molecule. The choice of the method used to generate the rigid molecule can be crucially important. Some of the search methods[6] are limited in treating only asymmetric units with Z`=1. The molecules in the three blind tests are shown in figure 5.1, together with the rate of success of the participants.

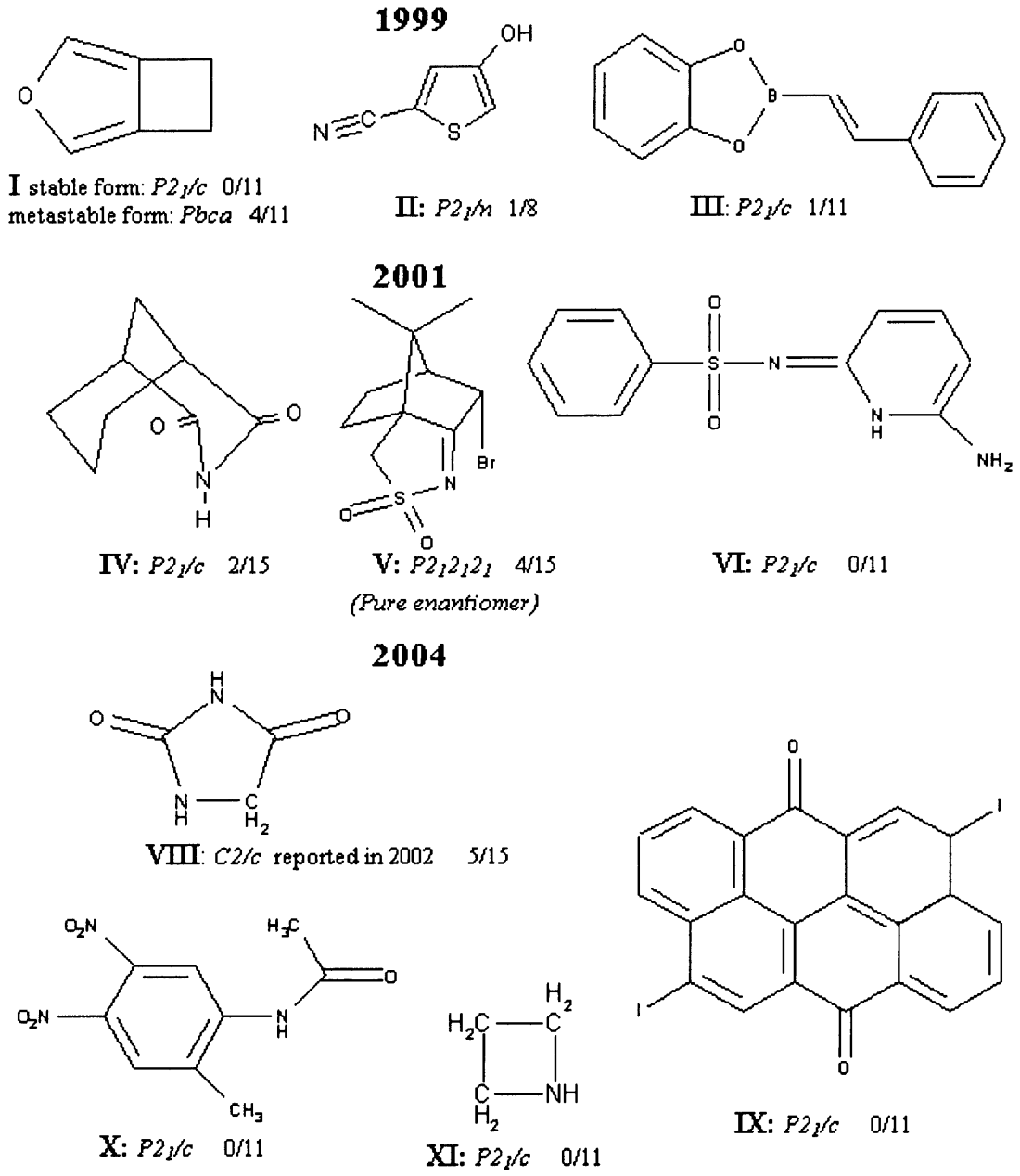


Fig. 5.1: Molecules studied in the three blind tests organised by the Cambridge CSD in 1999, 2001 and in 2004.

In the first blind test[2] molecule I had been previously crystallised in an orthorhombic phase (*Pbca*), but a break down in the cooling systems melted the crystal. Subsequently the compound was crystallised, this time in a monoclinic form ($P2_1/c$). Attempts to obtain the orthorhombic structure failed. A total number of 11 investigators participated to the first blind test, with every participant submitting 3 possible candidate predicted structures for each of the 3 molecules, for a total number of 99 possible solutions. However, the form of compound I (figure 5.1),which crystallised as first, subsequently transformed in another form, so that 4 crystallised structures in the blind test were finally considered, for

a total number of 132 possible solutions. The prediction of the structure of the first compound was successful for 4 of the 11 groups, who predicted the orthorhombic structure. Curiously if this structure had not appeared in the first crystallisation, the result would have been 0. Molecule II was successful predicted by 2 groups, whereas there was only one success in the prediction of the flexible molecule. Thus, the total rate of success for the first blind test is 7/44 or ~16%.

In the second blind test[3] in 2001, 17 groups participated to the prediction of the crystal structures of the three molecules (IV, V and VI in figure 5.1), submitting three choices per molecule, for a total of 51 possible correct answers. Two groups predicted correctly the crystal structure of the first molecule; four groups predicted the structure of second molecule, while no correct answers were supplied for the structure of the third flexible molecule. The rate of success in this case was 6/51 or ~ 12%. Hardly has been a general improvement with respect to the first test, but a significant improvement specifically in determining the crystal structure of rigid molecules was observed. As to the failure in the prediction of the highly flexible molecule, many participants noted the presence of an unusual distortion in the molecular conformation observed in the crystal, such that it was suggested that some hypothetical structures submitted in the blind test by the participants could be undiscovered polymorphs. Indeed another more stable form was later discovered[7], with a more usual hydrogen bond motif.

Here we report the study we performed for the compound XI of the third blind test in 2004.

5.1.2: Compound XI Blindtest 2004: Azetidine

The azetidine molecule was one of the four molecules (VIII, IX, X and XI in figure 5.1) investigated in the third blind test, which took place in 2004. As appropriately for a blind test[2-4] there were no crystal structures in the Crystal Structure Database (CSD)[1] for these molecules. The same rules as for the previous blind tests were followed.

The predicted structures were compared with experimental crystal structures, determined afterwards through XRD analysis to gather all the structural information related to the crystals. Figure 5.2 shows the structural formula of the azetidine:

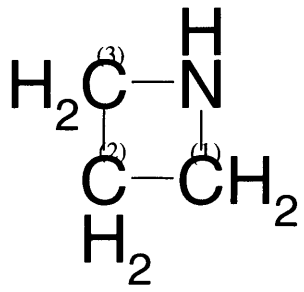


Fig. 5.2: Azetidine structural formula

Applications of this compound mainly involve its derivates, especially the carboxylic acids such as L-azetidine-2-carboxylic acid used as intermediate in the synthesis of polypeptides. From the computational point of view, the interest in this molecule is due to the fact that it is a small, relatively rigid molecule, whose characteristics make it a good candidate for a successful crystal structure prediction, as observed in the two previous blind-tests[2,3].

This molecule was additionally chosen for the third blind-test, because it was found that the crystal structure for one of the other molecules chosen (hydantoin) was already determined and had been presented in a poster session at a meeting of the British Crystallographic Association[8]!

5.2 Methods

5.2.1 Molecular Electronic Structure Calculation

Starting with the molecular structure drawn by the MOLDEN[9] program, we performed a geometry optimisation with the GAUSSIAN 98[10] program (see section 3.2). The wave function was calculated at the Hartree Fock/MP2 level[11], using the 6-31G** basis set. With the same method and the same basis set, we also calculated the electronic structure of the molecule and the electronic charge distribution, fitting the charges to the molecular electrostatic potential with the CHelpG algorithm[12](section 3.2). The program used for this calculation is again GAUSSIAN98[10]. We used the distributed multipole analysis (DMA[13]) up to hexadecapole order (section 1.4.2).

5.2.2 Crystal structure prediction and lattice energy minimisation

The molecule, optimised and calculated as described above, is the input file for the MOLPAK[6] program, which predicts all the possible crystal structures (see section 3.3).

We recall that the program treats only structures having one molecule per asymmetric unit cell. For each packing type, MOLPAK builds a maximum number of 200 structures and for each structure, after the construction of all the possible packing types, the program automatically calls the DMAREL[14] program to minimise the lattice energy of each packing type (see section 3.4). The inter-atomic potential used in the lattice energy minimisation is the *6-exp* (Buckingham) potential with the FIT parameters[15,16], the same as used for imidazole (chapter 4), whose parameters are reported in the table 3.2 (appendix 3.4 in chapter 3). The intra-molecular chemical bonds C-H and N-H were standardized, considering the typical values of 1.08 Å and 1.01 Å, respectively[17]. The charge-charge, charge-dipole, dipole-dipole electrostatic contributions to the lattice energy are evaluated by the Ewald summation technique[18]. All higher order terms in the multipole expansion are summed to a 15Å cut-off between the centres of mass of the molecule, which is also used for the repulsion-dispersion term. The final structures are tabulated, reporting for each structure, the packing type, the initial and final lattice energy, the initial and final volume of the unit cell, the lattice parameters of the unit cell and the lattice parameters of the reduced cell. For a given structure, we remove all of the similar structures, which have differences in lattice parameters less than specific threshold values, which distinguish two different structures, within the accuracy of the program:

ΔEnergy = 0.1 kJ mol^{-1}

ΔVol = 0.1 Å^3

Δ**a** (**b**,**c**) = 0.01Å

$\Delta\beta$ = 0.05 °.

For the predicted crystal structures we calculated the elastic constants and the free energy contribution through the second derivatives (calculated numerically) of the potential energy from the hessian matrix and estimating the phonon frequencies for **k=0**. These calculations were performed only for those structures in the lowest energy range, in order to have more criteria of choice for submitting the structures to the blind test. Dr. David Coombes[19] performed the calculation of the morphologic properties for the same set of structures, using the same intermolecular potential and fitted (ESP) electronic charges[12], instead of multipoles. Three of the predicted structures were submitted to the blind test[4], on the base of thermodynamic factors and morphologic properties.

Following the failure in the prediction, which will be reported and discussed later in the results, we minimised the experimental structure to test the quality of the FIT parameters

in the intermolecular potential with both the experimental molecular and the optimised geometry. In this latter case the further purpose was also to estimate the lattice energy and to compare this value with the lattice energy of the predicted crystal structures. Finally a morphologic study of the experimental structure was performed by Dr. David Coombes[19] in order to estimate the volume growth rate of the structure and to compare it with these of the predicted structures.

5.3 Results and Discussion

The molecular structure and geometry of azetidine was previously investigated in a gas-phase NMR study[20] and in a experimental microwave study and in a computational study[21]. These investigations showed that the molecule has two main independent vibrational motions: ring-puckering and N-H inversion (which proceeds via a transition state in which the N-C bonds are all coplanar). In particular the first makes the molecule flexible. The ring-puckering leads to two different conformations, one axial and one equatorial and *ab initio* studies[21] concluded that the axial conformation is unstable with respect to the equatorial. Geometry optimisation of the molecule confirms this result, giving the equatorial conformation as most stable geometry, as shown in figure 5.3.

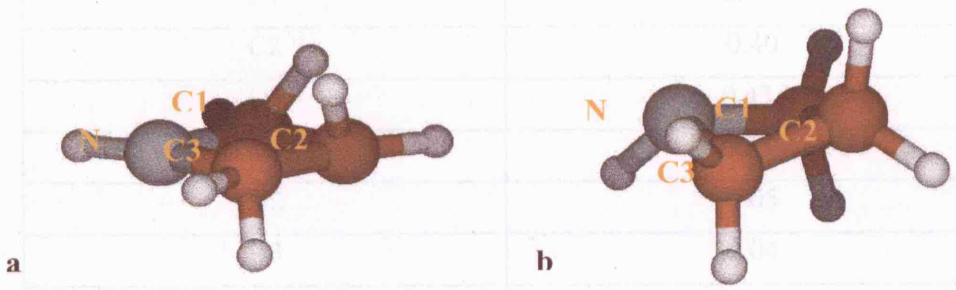


Fig. 5.3: a) Azetidine conformation before optimisation. **b)** azetidine "*equatorial*", the most stable conformation after optimisation.

Table 5.1 reports the results of the calculation of the molecular electronic structure of the optimised molecule, used in the subsequent CSP.

	E(MP2) /Hartree	*Dipole moment* /Debye	*C1N1C3* /o	*C1C3N1 H1* /o	*N-H Length* /Å	*Torsion angle* /o
Opt mol.	172.688	1.346	89.3	117.3	1.02	22.4

Table 5.1: Molecular properties of the optimised "gas phase" geometry of azetidine.

We note that the dipole moment is about half of that calculated in the imidazole (table 4.1 section 4.3.1), which is easily explained by the presence of only one N atom substituting a C atom in the ring. The N-H length is close to the standard value of 1.01 Å and the solid C1-C3-N1-H1 dihedral angle (table 5.1) indicates the bending angle of the bond N-H with respect to the molecular ring. Also for this molecule, as for imidazole, the charge density on the atoms (reported in the table 5.2) shows the presence of a substantial negative charge on the N atom and a positive charge on the H bonded to the N, which is in agreement with an electrostatic model of intermolecular (N)H···N-H hydrogen bond interaction.

Atom	*Charges (ChelpG)/ e*
N1	-0.78
C1	0.41
C2	-0.40
C3	0.43
H1	0.33
H2	-0.05
H3	-0.04
H4	0.08
H5	0.13
H6	-0.04
H7	-0.05

Table 5. 2: CHelpG atomic charges for the optimised azetidine. The atomic labels are as in figure 5.4.

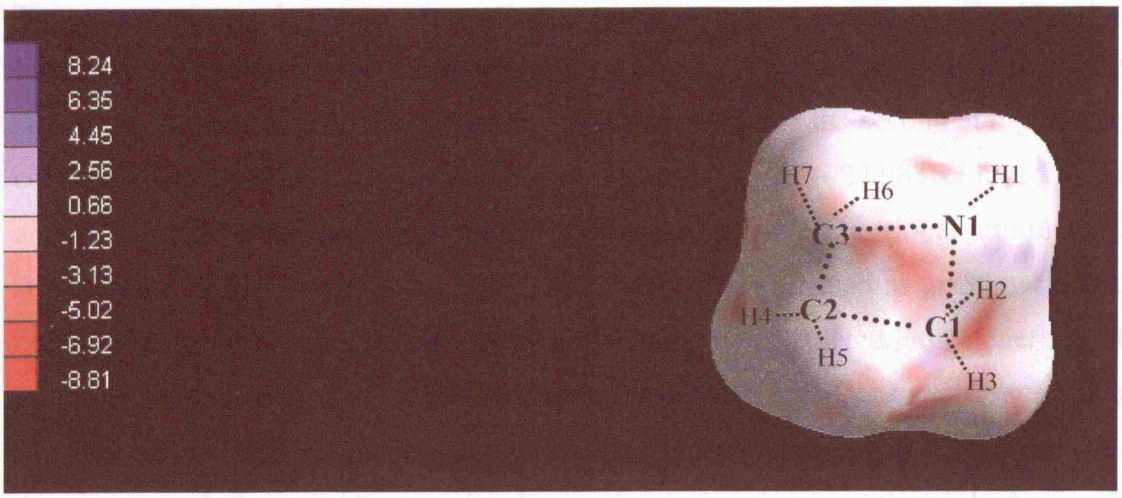


Fig. 5.4: CHelpG atomic charge density for the optimised azetidine. The electrostatic surface was generated through van der Waals spheres centred on each atom of the molecule. As shown in the legend of the colours, the negative charge is concentrated in the areas with red colour and the positive charge in the areas with blue colour.

The results obtained from the crystal structure prediction are shown in the following scatter-plot:

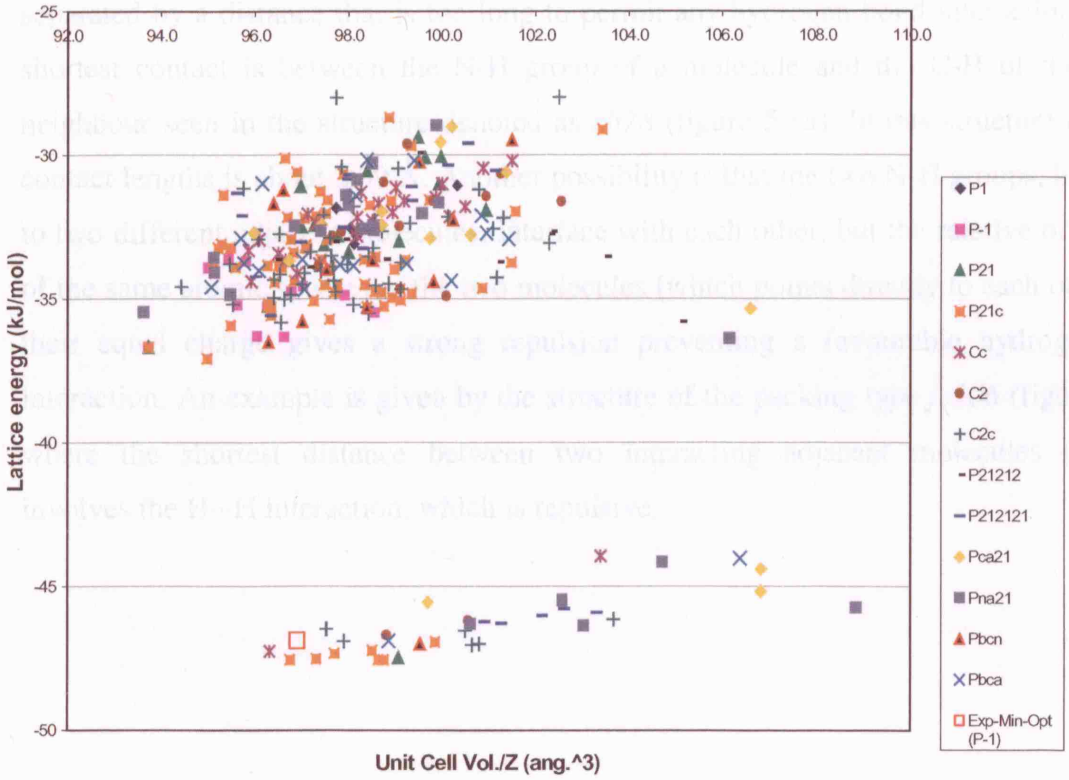


Fig. 5.5: Scatter-plot of the crystal structure prediction of azetidine crystal using the programs MOLPAK[6]+DMAREL[14].

The scatter plot in figure 5.5 shows two distinct groups of structures. The first, having lowest lattice energy, in the range between - 47.6 and - 43.9 kJ mol^{-1}, contains different structures, which are good candidates for crystallisation, and the second at higher energy, contains several structures in the range between - 37.073 and - 27.983 kJ mol^{-1}. A gap of about 7 kJ mol^{-1} in lattice energy separates these two groups. The global minimum (GM) structure has lattice energy equal to - 47.6kJ mol^{-1} and is denoted as *am85* packing type, corresponding to the space group $P2_1/n$. Contrasting the calculated lattice energies of the azetidine crystals with those of the imidazole crystals we note that the absolute values of the formers are much smaller and this is reasonable considering either the smaller dimensions of the azetidine molecule or the lack of conventional hydrogen bond acceptors with respect to imidazole.

Examining the patterns of the different predicted structures, we see that the discriminating factor between the two distinct groups is the intermolecular (N) H···N (H) hydrogen bond interaction, which is present in the low energy group of structures and absent in those belonging to the high energy group. In fact, considering the patterns of the highest lattice energy structures, we observe that the N-H groups of the molecules are separated by a distance that is too long to permit any hydrogen bond interaction and the shortest contact is between the N-H group of a molecule and the C-H of the closest neighbour seen in the structure denoted as *cb78* (figure 5.6a). In this structure the short contact lengths is about 2.39 Å. Another possibility is that the two N-H groups, belonging to two different adjacent molecules, interface with each other, but the relative orientation of the same atomic species in the two molecules (which points directly to each other) and their equal charge gives a strong repulsion preventing a favourable hydrogen bond interaction. An example is given by the structure of the packing type *fc106* (figure 5.6b), where the shortest distance between two interacting adjacent molecules (~2.05Å) involves the H···H interaction, which is repulsive.

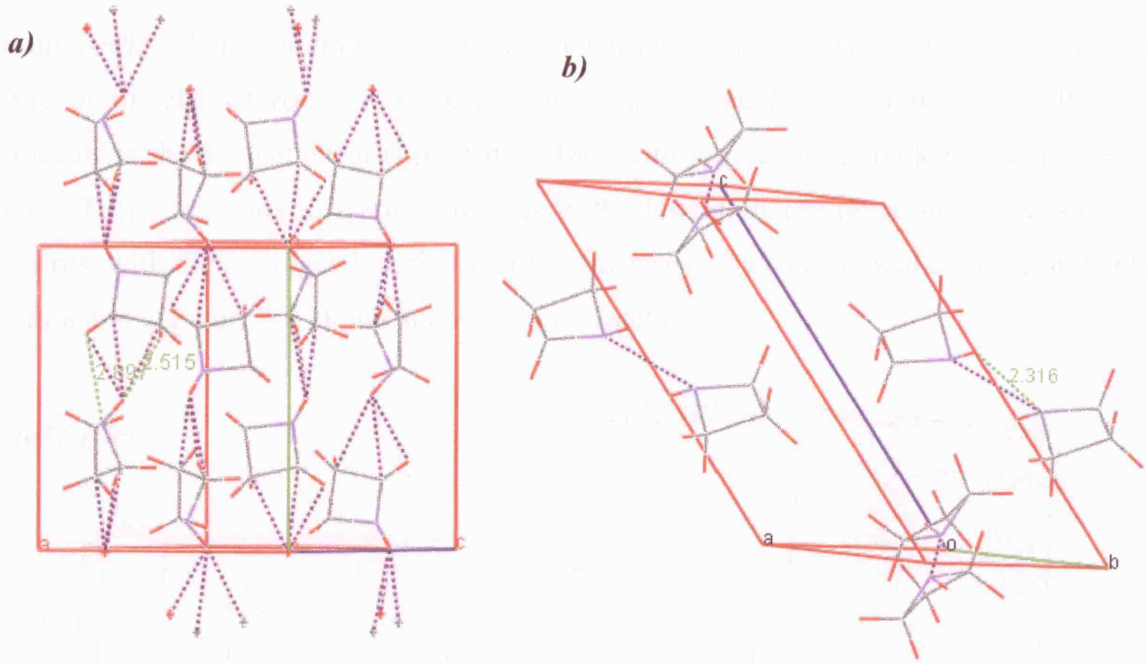


Fig. 5.6: Azetidine high-energy structure hydrogen bonds motifs. **a)** Packing type *cb78* (H-N groups don't point towards each other). **b)** Packing type *fc106* (H-N point to each other but the distance is too large for an hydrogen bond interaction). The hydrogen bond and short contacts are represented by the blue dotted lines.

In contrast all of the low-energy structures show an intermolecular hydrogen bond interaction.

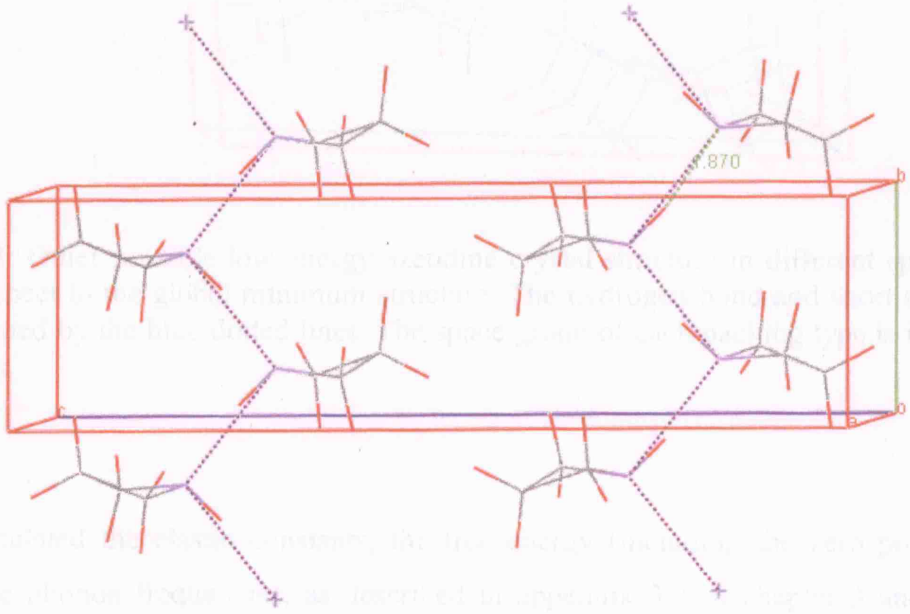


Fig. 5.7: Azetidine global minimum (*am85*) predicted structure. The hydrogen bond and short contacts are represented by the blue dotted lines.

In the *am85*, global minimum, structure (shown in figure 5.7) we observe an hydrogen bond motif, characterized by puckered chains along the ***b*** axis of the unit cell. The molecule in the parallel chain is rotated by 180° with respect to the molecule in the other chain. There are many different structures in the lowest energy range that show similar patterns, with the hydrogen bonding motif characterised by chains parallel to the unit cell axes, **a** and **c**. Figure 5.8 shows some of these structures.

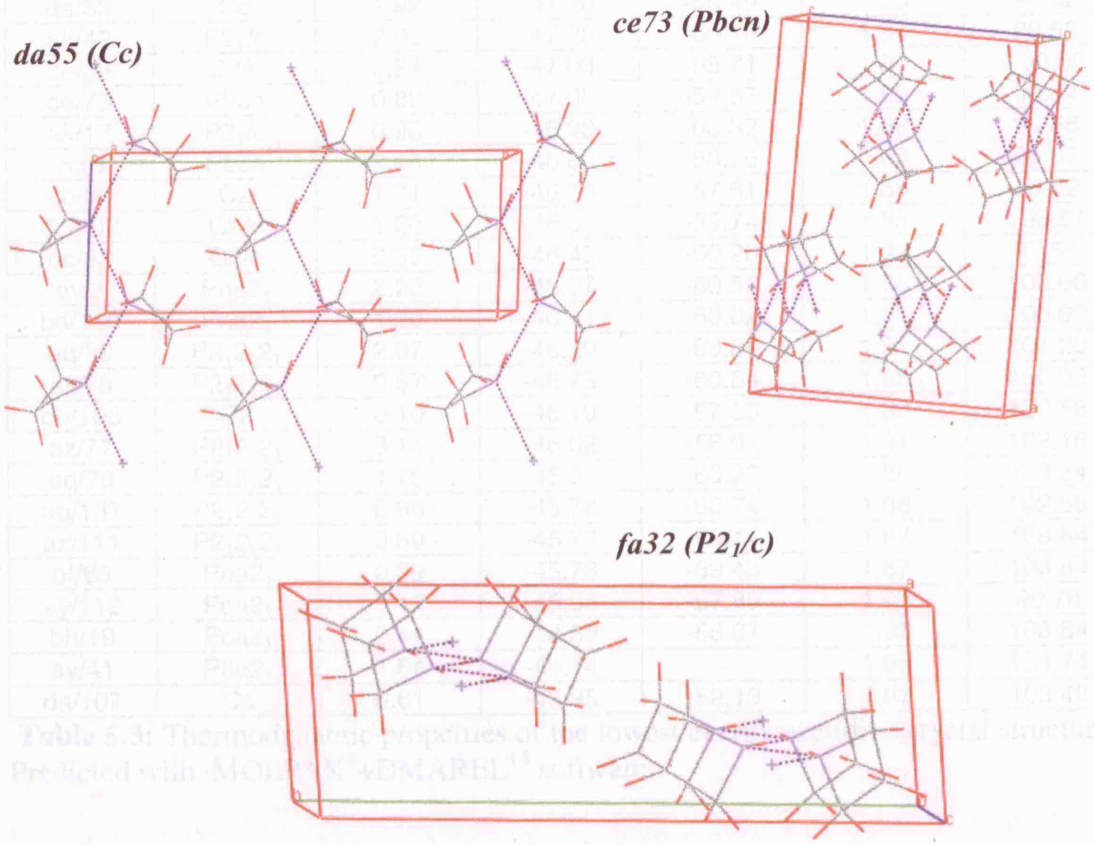


Fig. 5.8: Other possible low energy azetidine crystal structure in different space groups with respect to the global minimum structure. The hydrogen bond and short contacts are represented by the blue dotted lines. The space group of each packing type is indicated in brackets.

We calculated the elastic constants, the free energy (including the zero point energy) from the phonon frequencies, as described in appendix 3.2 of chapter 3 and, with the collaboration of Dr. David Coombes, the volume growth rate of the structures belonging to the lowest energy group in order to submit the three structures for the blind test. Table

5.3 reports all the data for the elastic constants, free energy, and hydrogen bond length of these structures.

Crystal Structure	*Space Group*	*Lowest Shear Constant (GPa)*	*Lattice Energy (kJ/mol)*	*Free Energy (kJ/mol)*	*HB length (Å)*	*Cell Vol per molec. (Å^3)*
am/85	$P2_1/n$	1.42	-47.56	-60.10	1.87	98.76
ak/48	$P2_1/c$	2.50	-47.56	-60.31	1.87	98.64
fa/32	$P2_1/c$	3.07	-47.54	-59.96	1.95	96.75
af/60	$P2_1$	1.53	-47.47	-59.48	1.87	99.08
da/55	Cc	1.92	-47.26	-56.49	1.96	96.32
ak/42	$P2_1/c$	2.02	-47.23	-60.19	1.88	98.52
de/33	C2/c	2.87	-47.03	-58.71	1.87	100.66
ce/73	Pbcn	0.80	-47.00	-59.57	1.92	99.54
ak/17	$P2_1/c$	0.98	-46.93	-60.32	1.88	99.86
cc/6	Pbca	2.58	-46.89	-58.75	1.93	98.87
db/98	C2	1.71	-46.70	-57.61	1.89	98.82
dc/102	C2/c	1.28	-46.52	-59.74	1.88	100.51
dc/33	C2/c	0.36	-46.48	-60.20	1.92	97.54
av/8	$Pna2_1$	2.20	-46.37	-60.53	1.86	103.06
bd/110	$Pna2_1$	1.20	-46.31	-60.02	1.88	100.62
aq/99	$P2_12_12_1$	2.67	-46.29	-60.84	1.91	101.29
az/78	$P2_12_12_1$	0.57	-46.23	-60.64	1.88	100.93
db/105	C2	0.10	-46.19	-57.53	1.88	100.58
az/77	$P2_12_12_1$	3.14	-46.02	-58.91	1.91	102.18
aq/79	$P2_12_12_1$	1.75	-45.91	-60.23	1.86	103.34
aq/137	$P2_12_12_1$	0.96	-45.76	-60.74	1.88	102.66
az/111	$P2_12_12_1$	0.89	-45.73	-59.44	1.87	108.84
bf/66	$Pna2_1$	0.89	-45.73	-59.43	1.87	108.84
ay/112	$Pca2_1$	0.12	-45.54	-57.89	1.97	99.70
bh/19	$Pca2_1$	0.44	-44.39	-56.87	1.92	106.84
av/41	$Pna2_1$	1.64	-44.14	-	1.95	104.74
da/107	Cc	0.61	-43.95	-52.13	1.97	103.42

Table 5.3: Thermodynamic properties of the lowest energy azetidine crystal structures Predicted with MOLPAK[6]+DMAREL[14] software.

Figure 5.9 shows the histogram of the volume growth rate for all these structures.

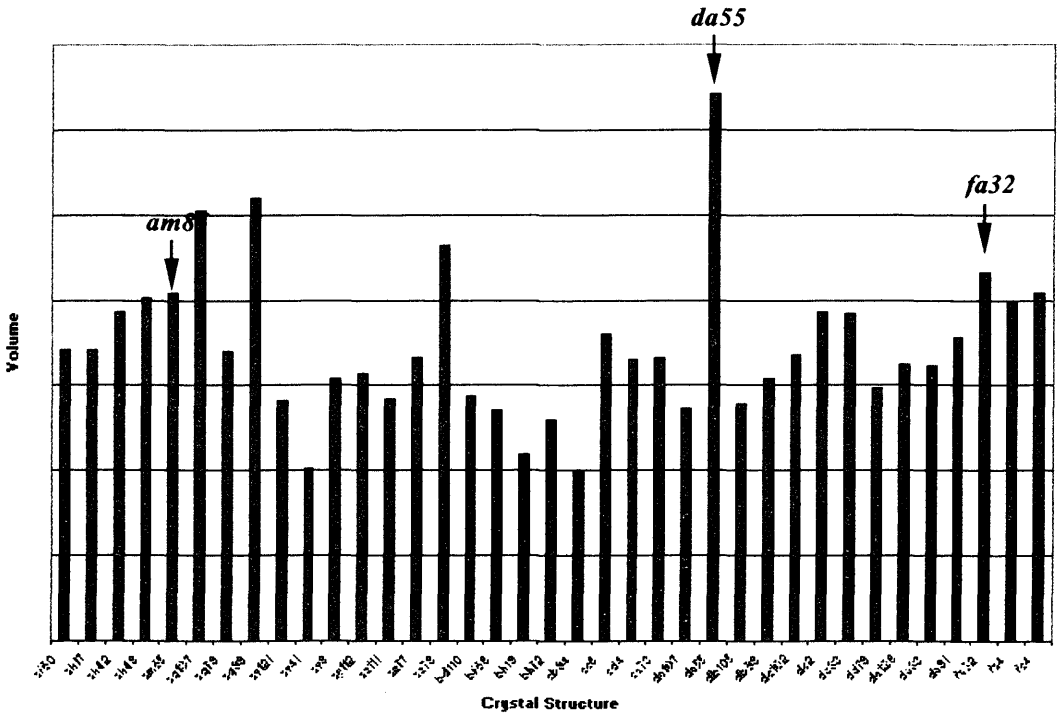


Fig. 5.9: Relative volume growth rate histogram for the azetidine crystals, calculated using the GULP program[22].

On the base of these thermodynamic factors and the approximate volume growth rate we submitted the following structures:

- *am85*: the global minimum in the lattice energy.
- *da55*: the highest volume growth rate.
- *fa32*: the second most densely packed structure.

The experimental structure, synthesized at 170K (with a R-factor = 0.05)[23] has a monoclinic structure in the $P2_1/c$ space group and unfortunately does not correspond to any of the structures predicted by MOLPAK[6]. This structure has two molecules in the asymmetric unit (Z`=2), which prevents the prediction of this structure with the software used[6], which generates structures with only one molecule per asymmetric unit. The two molecules have slightly different conformations and different energies and dipole moments, which also differ compared to the optimised molecule, as summarised in table 5.4.

	ΔE Hartree	*Dipole moment* Debye	*N5-H5 length* /Å	*C8-N5 length* /Å	*C6-N5 length* /Å	*C1C3N1H1* /o
Exp. Mol.1	0.003	1.373	1.010	1.465	1.474	117.5
Exp. Mol.2	0.003	1.420	1.010	1.465	1.470	116.7
Opt. Mol.	0	1.346	1.015	1.482	1.482	117.3

Table 5.4: Molecular properties for the two experimental molecules and the optimized structure. The atomic numbering is as in figure 5.10. $\Delta E = (E_{opt} - E_{exp})$.

The hydrogen – nitrogen bond length is longer in the optimized with respect to the experimental molecule. Moreover, the bending angle of the N-H group differs in the three molecules by a few degrees. In figure 5.10, we show the three molecules in their different geometries.

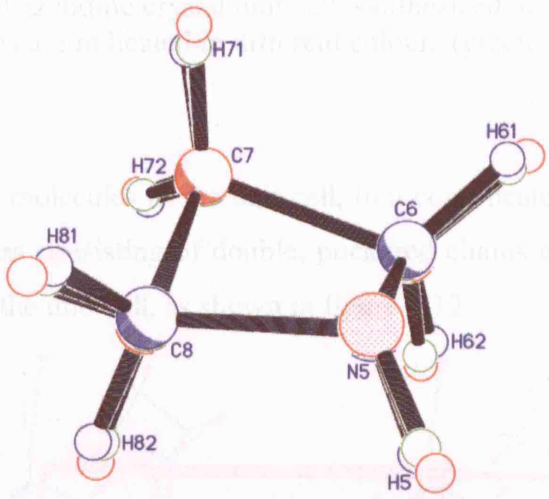


Fig. 5.10: The three molecular geometries of azetidine. Experimental molecule 1 (blue), experimental molecule 2 (yellow) and optimized molecule (red).

The two different experimental molecular geometries (shown in blue and green colour in figure 5.11) are atypical of such azetidines rings, since the lengths of the C-N bond (reported in table 5.4) are shorter than the average observed bond-lengths in other crystals of azetidines (1.484 Å; σ = 0.018 Å)[24]. This value is well reproduced in the optimized geometry of the molecule.

The hydrogen bond motif of the experimental crystal structure is shown in figure 5.11.

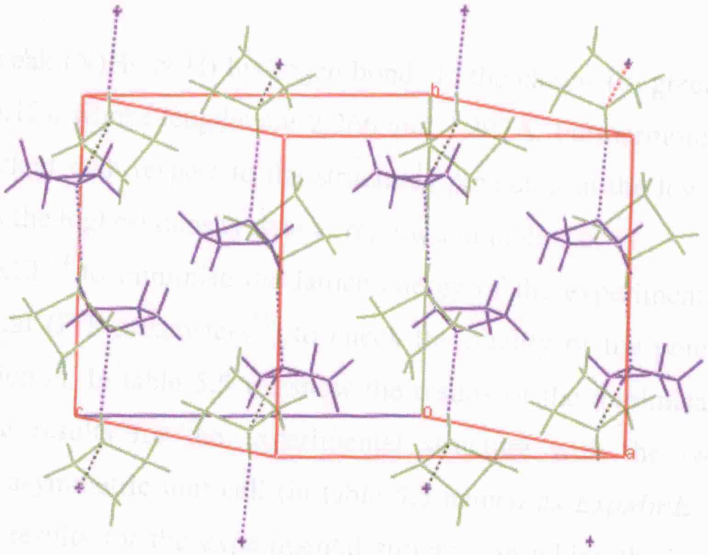


Fig. 5.11: Experimental azetidine crystal unit cell, synthesized at 170K. The two different molecular conformations are indicated in different colours (green and blue).

The structure has eight molecules in the unit cell, in a complicated hydrogen bond motif and may be described as consisting of double, puckered chains of molecules, which run parallel to the ***b*** axis of the unit cell, as shown in figure 5.12.

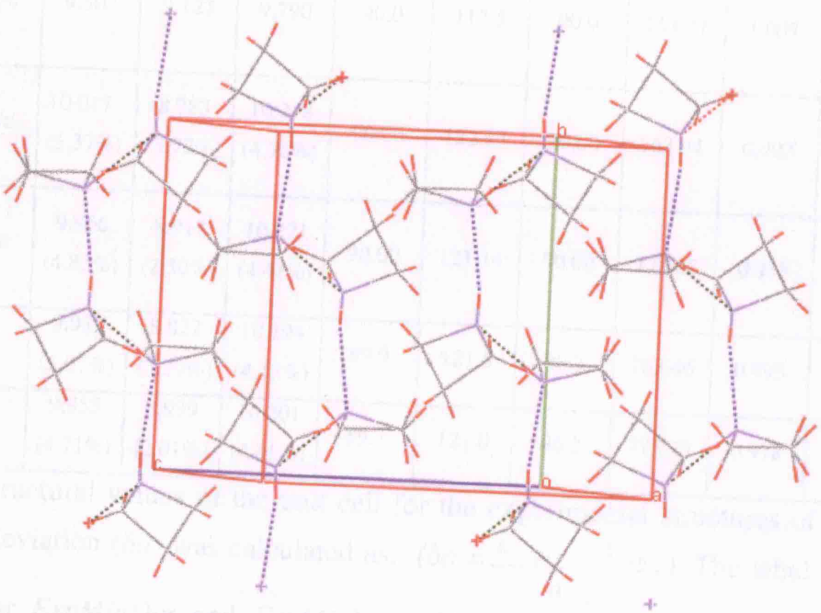


Fig. 5.12: Hydrogen bond motif of the experimental azetidine crystal. The two different hydrogen bonds are indicated with blue and green dotted lines.

There are two weak (N)H···N(H) hydrogen bonds in the chains (in green and blue dotted lines in figure 5.12), whose lengths are 2.266 and 2.297Å. Furthermore, this structure is very densely packed with respect to the structures predicted in the lowest lattice energy range, which has the highest density data as reported in table 5.5.

We used DMAREL[14] to minimise the lattice energy of the experimental structure with the *6-exp* potential (FIT parameters[15]) to check the quality of the potential parameters used in the prediction. In table 5.5 we show the results of the minimisation: the second row contains the results for the experimental structure with the two experimental molecules in the asymmetric unit cell (in table 5.5 named as *ExpMinExp*) and the third row contains the results for the experimental structure in which the two molecules are replaced by two molecules in the optimised geometry, used in crystal structure prediction for the blind test (in table 5.5 called as *ExpMinOpt).* For the intra-molecular chemical bonds C-H and N-H were standardized, using the typical values of 1.08 Å and 1.01 Å.

	Space Group	*a*/Å (δa)	*b*/Å (δb)	*c*/Å (δc)	α /°	β /°	γ /°	*Vol* /Å^3	*Density* /gcm^{-3}	*Lattice Energy*/ kJmol^{-1}
Exp.	$P2_1/c$	9.507	9.122	9.790	90.0	117.5	90.0	753.27	1.007	-45.48
ExpMin Exp (Meta)	$P2_1/c$	10.017 (5.37%)	8.783 (3.72%)	10.258 (4.78%)	90.00	122.17	90.00	763.94	0.993	-47.93
Exp MinOpt (Meta)	$P2_1/c$	9.966 (4.83%)	8.912 (2.30%)	10.221 (4.41%)	90.00	121.34	90.00	775.43	0.978	-46.84
ExpMin Exp	P-1	9.989 (5.07%)	8.822 (3.29%)	10.194 (4.13%)	89.9	121.5	86.3	763.46	0.993	-47.99
Exp MinOpt	P-1	9.955 (4.71%)	8.939 (2.01%)	10.201 (4.95%)	89.7	121.0	86.2	775.33	0.978	-46.88

Table 5.5: Structural values of the unit cell for the experimental structures of azetidine. The relative deviation (δa) was calculated as: ($\delta a = \frac{a_{exper} - a_{calc}}{a_{exp}}$). The label "*Meta*" in parenthesis for *ExpMinOpt* and *ExpMinExp* with symmetry $P2_1/c$, indicates that the minimized structure is not a proper minimum, but a metastable structure.

The table shows that for both molecules (in the experimental and optimized geometries) the crystal structure with a local minimum in lattice energy is in a different space group. The first optimization step brought the structure to a saddle point in the potential lattice energy surface (as shown by the presence of negative eigenvalue in one of the representation of the symmetry operators). This experimental crystal of azetidine is a metastable structure, corresponding to a "saddle" point in the lattice energy surface. The lattice energy value for this structure is - 47.93 kJ mol^{-1} for the structure with the two experimental molecular conformations (*ExpMinExp* (*Meta*) in table 5.5) and – 46.85 kJ mol^{-1} for that with the optimized molecular geometry (*ExpMinOpt*(*Meta*) in table 5.5). After we remove the third representation (with the negative eigenvalue), reducing the symmetry of the crystal from the space group $P2_1/c$ to $P\bar{1}$, minimization leads to a local minimum in the energy landscape with a value of -47.99 kJ mol^{-1} for the *ExpMinExp* (table 5.5) and - 46.88 kJ mol^{-1} for the *ExpMinOpt* (table 5.5). The difference between the two values is ~ – 0.05 kJ mol^{-1}, a small value, which shows the very smooth nature of the lattice energy surface in this system. The calculated zero point energy contribution for the experimental minimized structure (*ExpMinExp*) is ~ 3.85 kJ mol^{-1}, which is big enough to increase the energy of the crystal structure from the global minimum to the transition state and c.a.~ - 44 kJ mol^{-1}. In fact, the zero point energy of the crystal, at 0K, would overcome the barrier between the two $P\bar{1}$ structures, producing an averaged $P2_1/c$ structure. These thermodynamic considerations give more evidence of the metastability of the structure.

The RMS(%) deviations in the lattice parameters (described in the chapter about imidazole), are 5.08% and 3.98% respectively for the *ExpMinExp* and the *ExpMinOpt* structure of azetidine. This root mean square value is high for both minimizations, which indicates that the FIT parameters for the interatomic *6-exp* potential are inadequate for this system. In fact, the molecule has a large internal vibration inside the crystal structure, which is not well represented by this potential, which lacks intramolecular energy contributions. Figure 5.13 shows the difference between the three structures, discussed above.

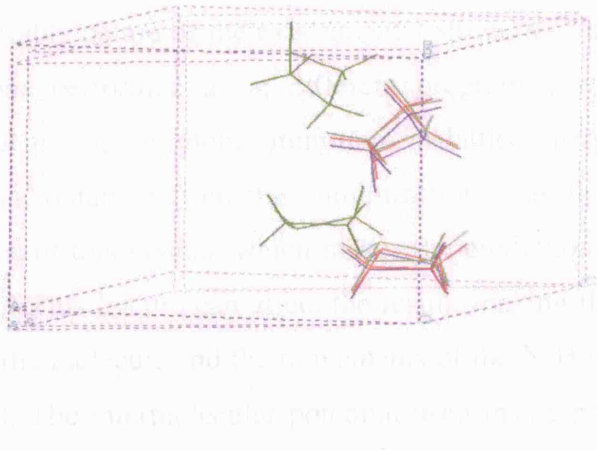


Fig. 5.13: Overlapped pictures of the three unit cells of the azetidine crystal, *Exp* (red), *ExpMinExp*(green), *ExpMinOpt* (blue).

5.4 Conclusions and further comments

From the geometry optimization it follows that the most stable conformation for azetidine molecule is the equatorial conformation, in agreement with previous studies[21]. It has some flexibility in the molecular ring and the presence of the N-H group in the molecule can give intermolecular hydrogen bonding interactions.

Crystal structure prediction using the MOLPAK[6]+DMAREL[14] procedure gives a lattice energy landscape, which is characterized by two distinctly separated groups of structures: a lowest energy and a higher energy group. Between these groups there is a difference in lattice energy of c.a. 7 kJ mol^{-1}. The key, discriminating factor involved in this distinct separation in energy is the (H)N···H(N) intermolecular hydrogen bond interaction. The hydrogen bond motif is characterised by puckered chains, which are differently oriented along the three axes of the unit cell. For the blind test, we submitted three structures, belonging to the lowest energy group of structures, on the base of thermodynamic and kinetics properties.

The presence of two molecules with slightly different geometries in the asymmetric unit cell of the experimental structure of azetidine crystal made the prediction of the azetidine crystal with the MOLPAK program impossible, since the latter predicts structures containing only one molecule per asymmetric unit cell. The use of other methods like some discussed in section 1.6.3 (Monte Carlo or Quasi-Monte Carlo methods) is necessary at this level.

The minimization of the experimental structure was particularly important first in testing the quality of the parameters used in the intermolecular potential and secondly in indicating the metastable nature of the experimental structure. This result was confirmed by other calculations performed using different programs[4], which could predict the crystal structure, but not as the global minimum in lattice energy. The third important piece of information, obtained from this minimization, was the smooth nature of the lattice energy surface of this system, which makes the prediction of the global minimum structure still harder. This feature can affect the result, making the system very sensitive to the flexibility of the molecule and the movements of the N-H group, even with the use of a better potential. The intermolecular potential used in the prediction is not accurate enough to permit a reliable prediction of azetidine's crystal structure, as shown in figure 5.10 and as indicated by the RMS(%) values reported in table 5.5 for the minimized structures, with the two molecules in the two different experimental configurations and with the optimized molecule in the unit cell. One possible reason for these difficulties is the high sensitivity of the potential to the presence of flexibility in the molecular ring of the azetidine and in the bending angle of the N-H group with respect to the ring, which can give dramatic changes in the (N)H···N(H) intermolecular hydrogen bond interaction in the crystal. Improvement in the treating flexible molecules during the lattice energy minimization procedure as well as improvements in including polarization effects in the potential might give better results even at 0K.

This molecule was expected to have those characteristics, which favored a successful prediction, but it turned out to be challenging as shown by other unsuccessful predictions performed by other participants to the blind test[24]. Even those who were able to generate crystal structures with Z`=2 found many distinct crystal structures in a small energy range above the global minimum, and were not able to predict the experimental structure as the global minimum. Furthermore by minimizing the lattice energy, in the post-analysis of this structure they found it to be metastable and the local minimum was obtained lowering the symmetry to a $P\bar{1}$, Z`= 4 crystal.

From these results, considering the potential surface of this molecule, we can conclude that the $P2_1/c$ experimental structure may be a thermal average over two symmetrical equivalent lower symmetry structures.

Improvements are necessary to develop a code, which can easily treat structures with more than one molecule in the asymmetric unit cell. Moreover we need to develop a

better model for the potential, which takes in account the intramolecular dynamics and vibrations, which can strongly contribute to the lattice energy.

Overall the third blind test[4] in 2004 has given less satisfactory results compared to the two previous blind tests. Among 18 participants and three candidate structures (the molecules are shown figure 1.18) none gave a successful for molecule IX, X and XI. Nevertheless it has to be noted that this blind test did not represent a "step back" in crystal structure prediction compared to the previous two blind tests. Indeed, the initial small rigid molecule of the candidate structures (molecule VIII in figure 5.1) had to be changed while the "competition" already started and many groups were calculating, or had already calculated the possible structures, as a conference report was discovered and hence the compounds rejected. Five of the 15 participants, who already produced the results from the crystal structure prediction, had predicted the experimental structure ranked as 1, 1, 1, 2 and 3 in the three possible choices. The azetidine molecule was given as new candidate compound for the blind test (molecule XI in figure 5.1). The prediction for this molecule was performed with less time compared to the other molecules in the blind tests.

This molecule resulted one of the most challenging system for crystal structure prediction[4]. Indeed, as already discussed above, the observed structure contains two molecules in the asymmetric unit, with slightly different torsion angles of the internal ring, forming puckered chains with space group $P2_1/c$. Some methodologies cannot deal with molecules with Z`>1, or avoided these systems, since a Z`>1 implies a much higher number of local minima, hence a more expensive search, which was an intrinsic limit for the search, but is not the only problem caused by this molecule. Indeed none of the groups, which could predict structures with Z` =2, was able to locate the structure in the "top 3". The closest prediction was given by Day[4], using a Monte Carlo stochastic search plus a lattice energy minimisation with the use of DMA for the electrostatic contribution. The structure was ranked as 4th in the choice. Other methods were able to predict the structure, but with much higher ranking. Some force fields also predict a planar configuration for the molecule, so that the experimental molecular conformation and hence crystal structure could not been modelled properly. The strained 4-membered ring is not common. All the searches performed on this system produce a large number of possible local minima, within few kJmol^{-1} (Pantelides & Karamertzanis[4] found 77 unique structures in 1 kJmol^{-1}!). Finally, a further investigation of the experimental structure indicated its energetic correspondence to a saddle point, such that the structure observed

may be the thermal average of two symmetrically equivalent lower symmetry structures. It has been recognised that many of polymorphic structures containing more than one molecule in the asymmetric unit cells are often "fossil relics"[25] of more stable and more symmetric structures with lower Z`, representing stages in a pathway towards more stable crystal structures, where the symmetry is higher[26]. These high Z` structures are often pseudo-symmetric, with the multipole moments in the asymmetric unit being related to each other by local symmetry elements[26]. Factors like irregular molecular shape, occurrence of small number of intermolecular interacting functional groups, frustration between the requirements of close packing and the satisfaction of strong directional intermolecular interactions between molecular polar groups, can lead the system to distort the packing arrangements in metastable structures, with higher number of molecules in the asymmetric unit cell (Z` > 1)[25]. Illustrative examples are given by triphenylgermane[25] and pentafluorophenol[26], characterised by having a polymorph with Z` > 1 and one at Z` = 1.

The failures in the reproduction of molecule IX are probably due to the problem of the molecular extension (large molecules with 35 atoms), as well as the lack of an accurate potential for the iodine species in the molecule. A simple isotropic spherical atom-atom potential was used in all the methods, although the halogens species are noted to have big anisotropy in the repulsive wall. Finally the lack of prediction of the flexible molecule X was mainly due to the absence of an accurate force field, which can model sufficiently well both the intra- and inter-molecular terms.

5.5 References

(1) Allen, F. H.; Motherwell, W. D. S. *Acta Cryst. Sec. B* **2002**, *58*, 407.

(2) Lommerse, J. P., et al. *Acta Cryst. Sec. B* **2000**, *56*, 697.

(3) Motherwell, S. W. D., Ammon, H. L. , Dunitz, J. D. , Dzyabchenko, A. m Erk, P. , Gavezzotti, A. , Hofmann, D. W. M. , Leusen, F. J. J. , Lommerse, J. P. M. , Mooij, W. T. M. ,Price, S. L. , Scheraga, H. ,Schweizer, B. , Schmidt, M. U. , van Eijck, B. P. ,Verwer, P. , Williams, D. E. *Acta Cryst. Sec. B* **2002**, *58*, 647.

(4) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. *Acta Cryst. Sec. B* **2005**, *61*, 511.

(5) van Eijck, B. P.; Kroon, J. *J. Comput. Chem.* **1999**, *20*, 799.

(6) Holden, J. R.; Du, Z. Y.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.

(7) Jetti, R. K. R.; Boese, R.; Sarma, J.; Reddy, L. S.; Vishweshwar, P.; Desiraju, G. R. *Angew. Chem.-Int. Edit.* **2003**, *42*, 1963.

(8) Yu, F.; Schwalbe, C. H. "Structure of hydantoin from a twinned crystal"; British Crystallographic Association Spring Meeting, 2002, Nottingham. England.

(9) Schaftenaar, G. MOLDEN; CAOS/CAMM Center Faculty of Science ; University of Nijmegen: P. O. Box 9010 6500 GL , Nijmegen, The Netherlands.

(10) Frisch, M. J. T., G.W.; Schlegel,H.N.; Scuderia,G.E.; Robb,M.A.; Chesseman,J.R.; Zakrzewski,V.G.; Montgomery,J.A.; Stratmann,R.E.; Burant,J.C.; Dapprich,S.; Millam,J.M.; Daniels,A.D.; Kudin,K.N.; Strain,M.C.; Farkas,O.; Tommasi,J.; Barone,V.; Cossi,M.; Cammi,R.; Mennucci,B.; Pomelli,C.; Adamo,C.; Clifford, S.; Ochterski,J.; Petersson,G.A.; Ayalla,P.Y.; Cui,Q.; Morokuma,K.; Malick,D.K.; Rabuck,A.D.; Raghavachari,K; Foresman,J.B.; Cioslowski,J.; Ortiz,J.V.; Stefanov,B.B.; Liu,G.; Liashenko,A.; Piskorz,P.; Komaromi,L.; Gomperts,R.; Martin,R.L.; Fox,D.J.; Keith,T.; Al-Laham,M.A.; Peng,C.Y.; Nanayakkara,A.; Gonzalez,C.; Challacombe,M.; Gill,P.M.W.; Johnson,B.G.; Chen,W.; Wong,M.W.; Andres,J.L.; Head-Gordon,M.; Replogle,E.S.; Pople,J.A.; *GAUSSIAN 98*; Gaussian Inc.: Pittsburgh, 1998.

(11) Moeller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(12) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(13) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(14) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628.

(15) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

(16) Williams, D. E.; Cox, S. R. *Acta Cryst. Sec. B* **1984**, *40*, 404.

(17) Allen, F. H. K., O. ; Watson, D. G. ; Brammer, L. ; Orpen, A., G. *J. Chem. Soc., Perkin Trans.* **1987**, *2*, S1.

(18) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.

(19) Coombes, D. S.; Catlow, C. R. A.; Gale, J. D.; Hardy, M. J.; Saunders, M. R. *J. Pharm. Sci.* **2002**, *91*, 1652.

(20) Friedman, B. R., Chauvel Jr., J. P. , True, N. S. *J. Am. Chem. Soc.* **1984**, *106*, 7638.

(21) Lopez, J. C., Blanco, S. , Lesarri, A. , Alonso, J. L. *J. Chem. Phys.* **2001**, *114*, 2237.
(22) Gale, J. D.; Rohl, A. *Mol. Simul.* **2003**, *29*, 291.
(23) Parsons, S., *Univeristy of Edinburgh, Chemistry Department.*
(24) Day, G., et al. **2004**.
(25) Steed, J. W. *CrystEngComm* **2003**, *5*, 169.
(26) Das, D.; Banerjie, R.; Mondal, R.; Howard, J. A. K.; Boese, R.; Desiraju, G. R. *Chem. Commun.* **2006**, 555.

Chapter 6

CYCLOPENTANE CRYSTAL STRUCTURE PREDICTION

6.1 Introduction

Cyclopentane is an organic compound which is used in the preparation of synthetic resins and rubber adhesive, and it has a growing application as foam blowing agent in the refrigeration industry, substituting the environmentally unfriendly chlorofluorocarbons[1].

Cyclopentane molecule is an alicyclic hydrocarbon with chemical formula C_5H_{10}, which consists of a ring of five carbon atoms with two hydrogen atoms for each carbon atom, located above and below the plane of the ring. The molecular structure, which has two thermodynamically stable conformations, known as "*half-chair*" and "*envelope*"(figure 6.1), is characterised by a high flexibility in the molecular ring.

envelope *half-chair*

Fig. 6.1: The two thermodynamically stable cyclopentane molecular conformations.

From an electronic point of view this molecule does not show the presence of polar groups, such as N-H or O-H, so that no predominant intermolecular strong electrostatic interactions will be present in its solid state.

Cyclopentane compound has boiling point at 322K and melting point at 180K, while its solid state is characterised by the presence of three different phases. Plastic phase I has a hexagonal unit cell[2], for which it was not possible to solve the X-ray diffraction data of

the structure. A second plastic phase, known as phase II, has not been possible to solve, because of the presence of a high disorder in the hydrogen motion[2,3] and of an orientational dynamical disorder, which was also observed in plastic phase I. At low temperature (T < 120K, P=1bar) a third ordered monoclinic crystalline phase (phase III) has been observed and solved at 93K by R.Boese[4], which determined the following lattice parameters by using single crystal X-ray diffraction analysis:

a = 10.03 Å, b = 5.335 Å, c = 9.61Å, α = 90.0, β = 113.2, γ = 90.0

Figure 2 shows the phase diagram of cyclopentane.

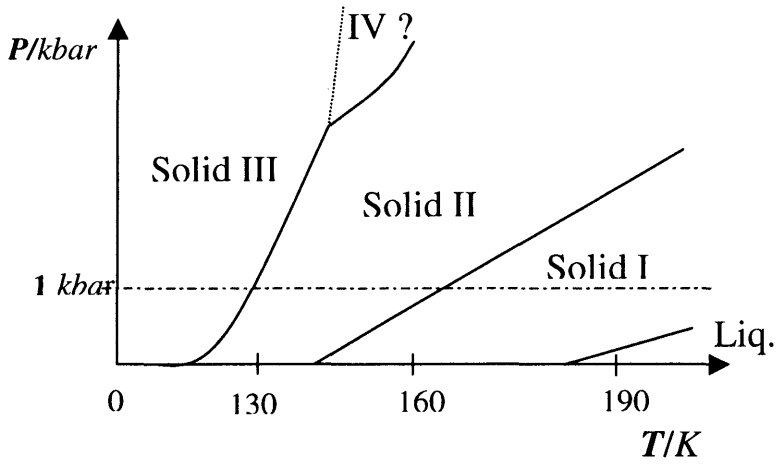


Fig. 6.2: Phase diagram of cyclopentane. The notation (IV?) wants to indicate the fourth solid phase at high pressure, whose existence has been hypothesized by Webster and Hoch[5], but never observed experimental evidence in XRD diffractions.

6.2 Methods

The geometry optimisation of the molecule and its electronic structure calculation were performed through *ab initio* calculation, using the program GAUSSIAN98[6]. The method used to calculate the molecular wave function is MP2[7] with basis set 6-31G**. The hydrogen bonds were set to the standard neutron value of 1.01Å.

The crystal structure prediction follows the usual steps, which are reported in figure 1.9 in chapter 1, and was performed using the MOLPAK[8] program and for each of the generated crystal structures was performed a lattice energy minimisation using the DMAREL[9] program. The potential used is a *6-exp* form (section 1.5.2), with FIT parameters[10], for the repulsive-dispersive contribution to the lattice energy, plus an electrostatic term, modelled with a distributed multipole analysis (DMA[11]), up to the hexadecapole order, of the atomic point charges, which were derived from *ab initio* calculations, using the CHelpG algorithm[12]. The multipole expansion is treated with

Ewald summation[13] with a cut-off of 15Å, which is the same of that used for the repulsive-dispersive contribution in the atom-atom interaction. Finally the generated structures are summarised in a table, which reports for each structure the packing type, the initial and final lattice energy, the initial and final volume of the unit cell, the lattice parameters of the unit cell and the lattice parameters of the reduced cell. For a given structure, we remove all of the similar structures, which have differences in lattice parameters less than specific threshold values, which distinguish two different structures, within the accuracy of the program:

ΔEnergy = 0.1 kJ mol^{-1}

ΔVol = 0.1 Å^3

Δa (b,c) = 0.01Å

$\Delta\beta$ = 0.05 °.

6.3 Results and Discussion

The *ab initio* optmisation of the molecule indicates the "*half-chair*" conformation to be the most stable in the isolated "gas phase", which is shown in figure 6.2:

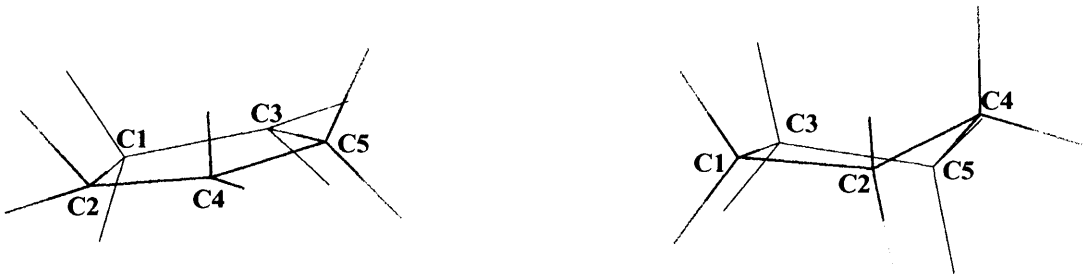


Fig. 6.3: Optimisation of the molecular conformation of cyclopentan. **a)** planar intial molecule. **b)** "*half-chair*" final gas phase molecule.

Table 6.1 shows the energy values, the dipole moments and the geometrical parameters for the molecule in the "*gas phase*" conformation and the molecule in the experimental conformation of the monoclinic phase.

	Experimental molec	Optimised molec (gas phase)
TorsionalAngle (C2-C1-C3-C5) /Degree	14.44	18.57
H-C bond length /Å	0.96 - 1.04	1.09
Dipole moment /Debye	0.069	0.013
Energy /Hartree	-195.827	-195.896

Table 6.1: Geometrical, thermodynamical and electrical properties of cyclopentane molecule, in the gas phase and in the experimental conformation observed in monoclinic phase III.

The molecular geometries have very small dipole moments, which makes assume the presence of weak intermolecular interactions and the lack of preferred directional electrostatic interactions, in leading to the crystal formation. Figure 6.3 shows the two molecular conformations, reported in table1.

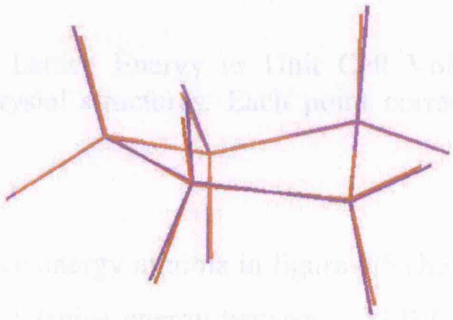


Fig.6.4: Difference between the experimental (monoclinic crystal) molecular conformation (orange) and the "*gas phase*" conformation (blue) of the cyclopentane.

There are small differences in the physical and geometrical parameters in the two different conformations, resulting in a shorter bond length of the group H-C and a smaller torsional angle in the experimental conformation.

Figure 6.4 shows the plot of the lattice energy vs volume/molecule for each of the generated structures in order to give a picture of the "lattice energy landscape" for the system, using the "*gas phase*" optimised molecular conformation.

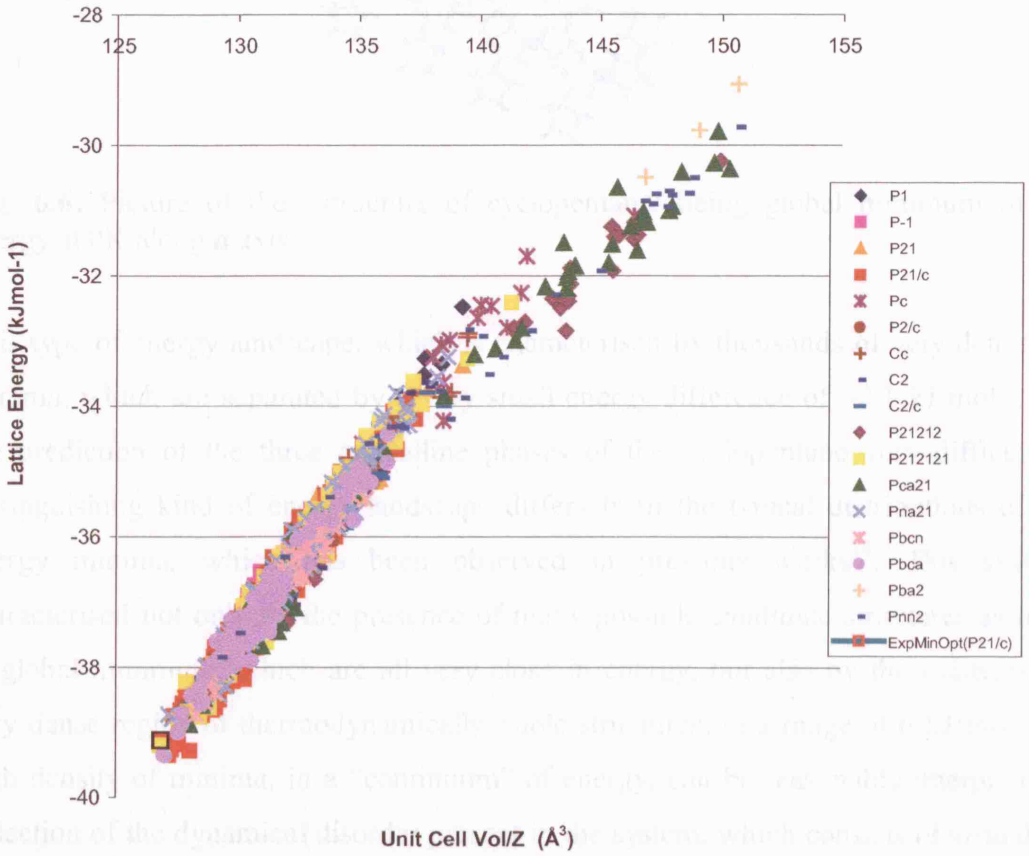


Fig. 6.5: Scatter plot of Lattice Energy *vs* Unit Cell Volume per molecule, for the cyclopentane molecular crystal structures. Each point corresponds to a specific crystal structure.

The scatter-plot of the lattice energy minima in figure 6.5 shows a continuous presence of local minima in a range of lattice energy between -33kJ mol^{-1} and -39kJ mol^{-1}. The global minimum is at -39.361 kJ mol^{-1}, corresponding to a structure with space group Pbca, shown in figure 6.6.

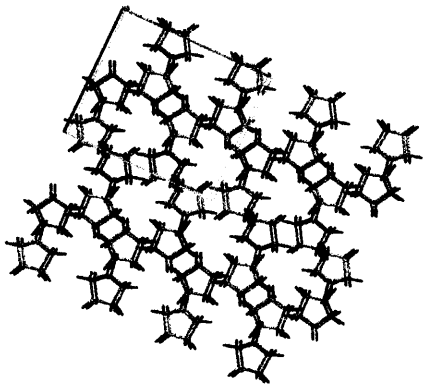


Fig. 6.6: Picture of the structure of cyclopentane, being global minimum in lattice energy at 0K along ***a*** axis.

This type of energy landscape, which is characterised by thousands of very densed local minima, which are separated by a very small energy difference of ~0.1 kJ mol^{-1}, makes the prediction of the three crystalline phases of the cyclopentane very difficult. This distinguishing kind of energy landscape differs from the typical distributions of lattice energy minima, which has been observed in previous works[14]. This system is characterised not only by the presence of many possible candidate structures as minima as global minimum, which are all very close in energy, but also by the existence of a very dense region of thermodynamically stable structures, in a range of 6 kJ mol^{-1}. Such high density of minima, in a "continuum" of energy, can be reasonably interpreted as a reflection of the dynamical disorder present in the system, which consists of virtually free rotational motion in the cyclopentane molecules. Because of the lack of any electrostatic strong interaction, a small change in the molecular hydrogen orientations does not strongly affect the lattice energy of the crystal.

Figure 6.7 shows an enlargement of the scatter plot in the range of about 1 $kJmol^{-1}$ from the global minimum.

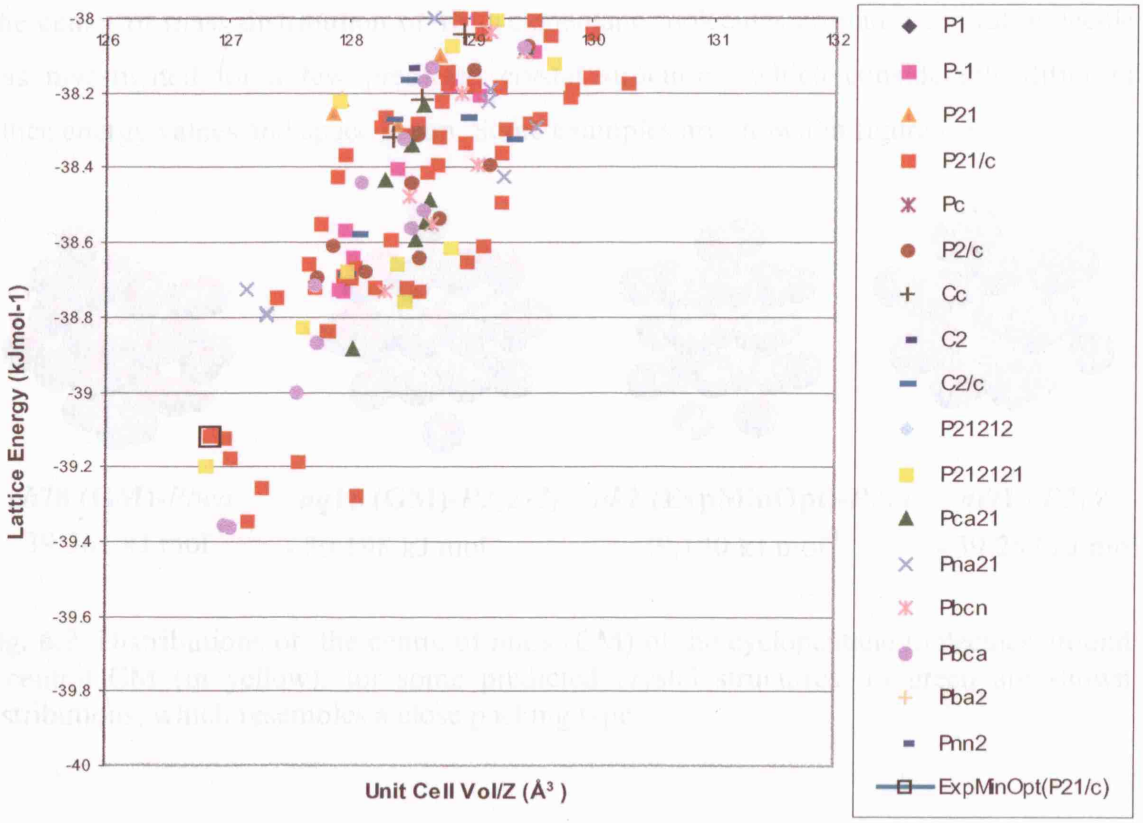


Fig. 6.7: Enlargement of the scatter plot for the cyclopentane molecular crystal structures in a range of about 1kJ/mol.

Considering the high rotational motion and flexibility of cyclopentane molecules in this system, we investigated the possible presence of a closed packed distribution of the molecular centre of masses in the predicted structures, similar to the close packing spheres distributions, hexagonal closed packing (hcp) and cubic closed packing (ccp), which are characterised by a number of 12 first neighbours around a central sphere (figure 6.7) and are typically observed in compact ionic crystals (i.e. NaCl)[15].

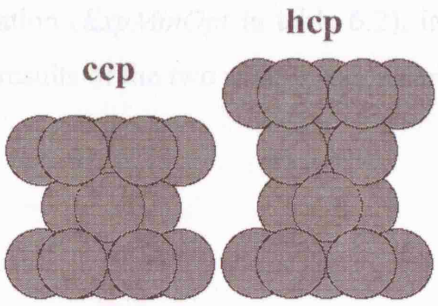
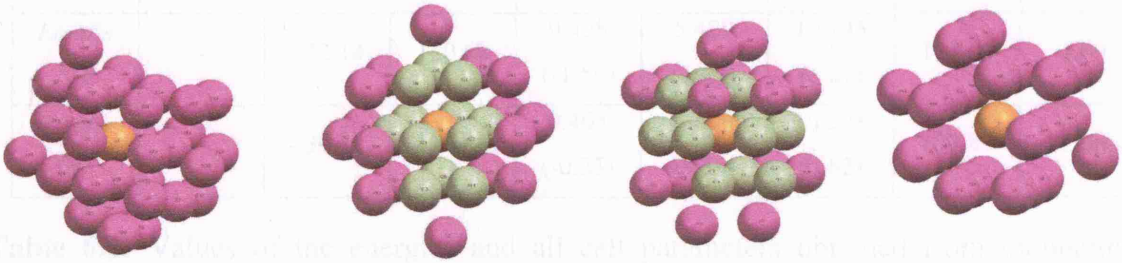


Fig. 6.8: Typical ccp and hcp closed packing spheres distribution.

The center of mass distribution of 18 cyclopentane molecules around a central molecule was investigated for a few predicted crystal structures, which considerably differ in lattice energy values and space group. Some examples are shown in figure 6.8.



cb*78 (GM)-*Pbca
- 39.361 kJ mol^{-1}

***aq*18 (GM)-$P2_12_12_1$**
- 39.198 kJ mol^{-1}

***ak*2 (ExpMinOpt)-$P2_1/c$**
- 39.120 kJ mol^{-1}

***ai*91 –$P2_1/c$**
- 39.257 kJ mol^{-1}

Fig. 6.9: Distributions of the centre of mass (CM) of the cyclopentane molecules around a central CM (in yellow), for some predicted crystal structures. In green are shown distributions, which resembles a close packing type.

Many predicted structures have very similar centres of mass distributions: some of the latter resemble a closed packed spheres model, whilst others are far from being close packed. The main distance of the first neighbours from the central centre of mass is ~ 5Å, so that, on the average, the distributions seem not to have a closed packed form.

A lattice energy minimisation using DMAREL[9] program was performed on the experimental structure of the monoclinic phase of cyclopentane, using the same potential, which has been used in CSP. Initially we minimised the lattice energy of the crystal structure containing the experimental molecular conformation (*ExpMinExp* in table 6.2), in order to test the quality of the potential used for this system. Afterwards we minimised the lattice energy of the experimental monoclinic crystal structure, with the molecule in the "*gas phase*" conformation (*ExpMinOpt* in table 6.2), in order to compare it with the predicted structures. The results of the two minimisations are reported in table 6.2.

	Method	*Latt.En.* */kJmol*$^{-1}$	*Vol/Z* */Å*^{3}*mol*$^{-1}$	*a* */Å* (*δ*%)	*b* */Å* (*δ*%)	*c* */Å* (*δ*%)	*β* /°	*RMS%*
Exp.	Neutron T = 93K	- 41.62	117.54	9.578 -	5.328 -	10.018 -	113.12	-
ExpMin Exp	-	- 42.14	120.68	9.428 (-1.56)	5.429 (1.90)	10.145 (1.27)	111.63	1.60
ExpMin Opt	-	- 39.12	126.86	9.405 (-0.25)	5.619 (3.50)	10.208 (0.62)	109.83	2.06

Table 6.2: Values of the energies and all cell parameters obtained from monoclinic crystal minimisation using DMAREL[9]. The root mean square coefficient (*RMS%*) were calculated as: $RMS\ (\%) = \sqrt{\frac{(\delta a)^2 + (\delta b)^2 + (\delta c)^2}{3}}$, where δa (and similarly δb and δc) has the expression: $\delta a = 100 \cdot \frac{(a_{exp\,er.} - a_{calc.})}{a_{calc.}}$.

The *RMS(%)* values for the cyclopentane indicate the good quality of the intermolecular potential in relaxing the crystal structure, either for the experimental or for the optimised "*gas-phase*" conformation of the molecule in the unit cell.

The energy minimised monoclinic phase of cyclopentane, with the optimised molecule in the unit cell, corresponds with one of the predicted structures (*ak2* packing type), but the latter does not correspond to the global minimum in the lattice energy, differing by a tiny energy difference of 0.2 kJ mol^{-1}.

Table 6.3 reports the cell parameters of the reduced cell (calculated using the program PLATON[16]), which confirm the agreement between the two structures.

Reduced Cell	*a* */Å*	*b* */Å*	*c* */Å*	*α* /°	*Latt. En* */kJmol*$^{-1}$.	*Vol/Z* */Å*^{3}
ExpMinOpt	5.619	9.405	10.208	109.83	-39.120	126.86
ak2	5.619	9.407	10.208	109.83	-39.120	126.86

Table 6.3: Reduced cell parameters for the relaxed experimental monoclinic structure of cyclopentane with optimised molecular conformation in the unit cell (*ExpMinOpt*) and the predicted structure *ak2*.

The agreement between the experimental energy minimised (*ExpMinOpt*) and the predicted structure (*ak2*, figure 6.10), has been assessed by calculating the RMS of two overlapped coordination spheres of 15 molecules corresponding to these two structures,

using the program COMPACK[17] (see section 4.4.1), with a tolerance of 15% in the intermolecular distance. A RMS value of 0.0014Å confirms the very good agreement between the two structures, which has also been noticed between the two X-ray powder patterns the two structures, calculated with the Mercury[18] program, with λ = 1.54056 Å (figure 6.10).

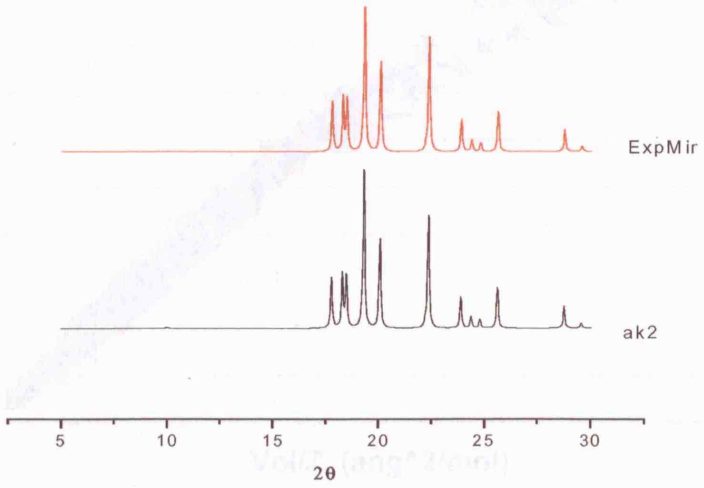


Fig. 6.10: Simulated X-ray powder patterns for *ExpMinOpt* crystal structure of the monoclinic phase of cyclopentane (red) and *ak2* (black) predicted structure. The radiation wavelength is set up at λ = 1.54056 Å.

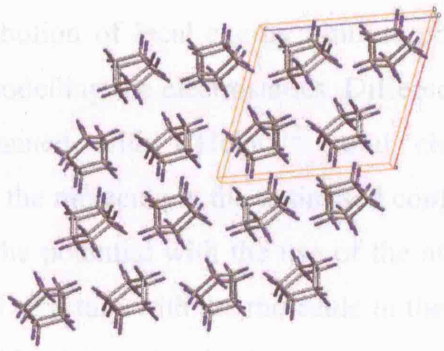


Fig. 6.11: Cyclopentane relaxed monoclinic structure, with cyclopentane molecule in "*gas phase*" conformation (*ExpMinOpt*) along ***b*** axis.

Because of the lack of polar groups in the molecule and the consequent absence of strong preferred electrostatic interactions in the molecular crystal, we assumed that the solely use of atomic point charges could equally predict a lattice energy landscape similar to that obtained with DMA[11]. Hence, we performed a CSP using the same repulsion-

dispersion potential but using the *ab initio* calculated atomic point charges, and the predicted lattice energy landscape is reported in figure 6.11.

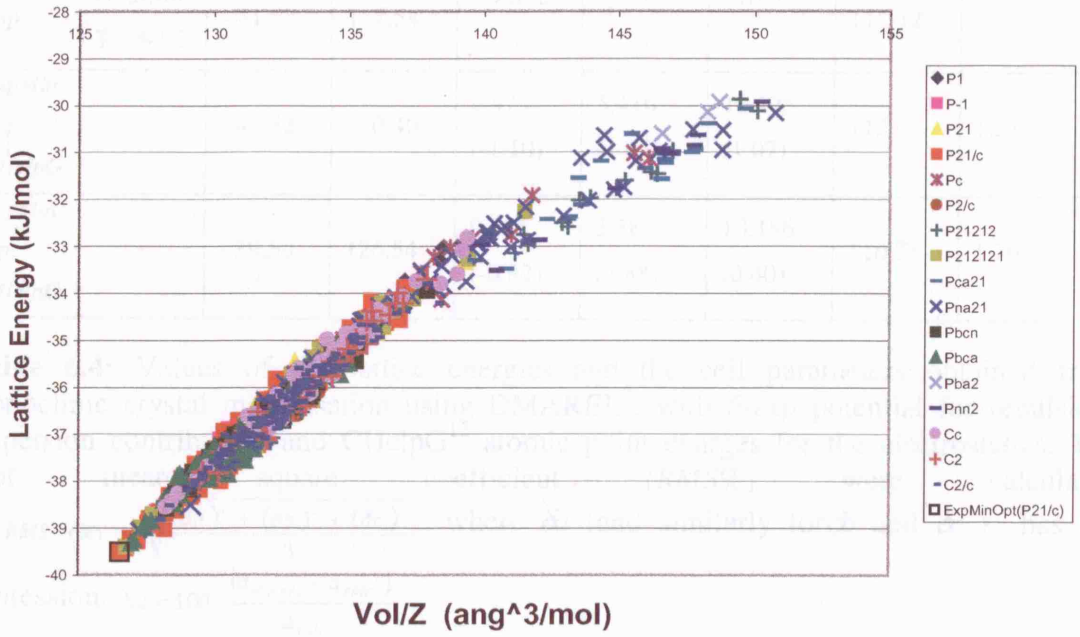


Fig. 6.12: Lattice energy landscape for the cyclopentane using *6-exp* potential (Buckingham potential), with FIT parameters[10], for the repulsive-dispersive contribution and CHelpG[12]point charges for the electrostatics contribution.

Figure 6.12 shows a distribution of local energy minima very similar to that observed with the use of DMA, in modelling the electrostatics. Differently from CSP with DMA[11], the global minimum obtained with CHelpG[12] point charges corresponds to the experimental structure with the molecule in the optimised conformation (*ExpMinOpt*).

To evaluate the quality of the potential with the use of the atomic point charges we also minimised the experimental structure with the molecule in the experimental conformation (*ExpMinExp*) and that with the molecule in the *ab-initio* optimised conformation (*ExpMinOpt*), as we previously did to test the potential with the DMA[11]. The results are reported in table 6.4.

	Method	*Latt.En.* */kJmol^{-1}*	*Vol/Z* */Å^{3}mol^{-1}*	*a* */Å* *(δ%)*	*b* */Å* *(δ%)*	*c* */Å* *(δ%)*	*β* */°*	*RMS%*
Exp.	Neutron T = 93K	- 41.62	117.54	9.578 -	5.328 -	10.018 -	113.12	-
ExpMin Exp ***CHelpG***	-	- 42.52	120.40	9.473 (-1.10)	5.416 (1.65)	10.125 (1.07)	112.01	1.31
ExpMin Opt ***CHelpG***	-	- 39.50	126.54	9.515 (-0.92)	5.585 (2.88)	10.186 (0.40)	110.75	1.76

Table 6.4: Values of the lattice energies and the cell parameters obtained from monoclinic crystal minimisation using DMAREL[9], with *6-exp* potential for repulsion-dispersion contribution and CHelpG[12] atomic point charges for the electrostatics. The root mean square coefficient (*RMS%*) were calculated as: $RMS\ (\%) = \sqrt{\frac{(\delta a)^2 + (\delta b)^2 + (\delta c)^2}{3}}$, where δa (and similarly for δb and δc) has the expression: $\delta a = 100 \cdot \frac{(a_{exp\,er.} - a_{calc.})}{a_{calc.}}$.

From the data in the table 6.4 we see that, the use of the atomic point charges, in modelling the electrostatic contribution, does not affect the good quality of the potential. On the contrary the RMS(%) values obtained from the minimisation of the experimental cyclopentane crystal structures, using CHelpG point charges (table 6.4) are even lower than those obtained in the energy minimisation, using DMA[11] (table 6.2).

6.4 Conclusions

CSP of cyclopentane using a method based on the sole static lattice energy minimisation, within a rigid molecular model, shows a peculiar energy landscape, which is characterised by a "continuum" of minima, which makes the prediction of the three known phases of cyclopentane very difficult. Such a characteristic distribution of minima, which was obtained by modelling the electrostatics either with a multipole expansion of the atomic charges (DMA[11]) or with CHelpG atomic point charges, reflects the presence of dynamical disorder, which has been experimentally observed in two of the cyclopentane solid phases[2,3], and causes problems in solving the X-ray diffraction patterns of these two structures[2]. Indeed this kind of energy landscape implies the presence of a local minimum for many different orientations of a rigid cyclopentane molecule in the unit cell, which differ of few degrees from each other, and this behaviour

can be associated with a molecule, which can undergo a multitude of different orientations in the crystal structure and this might be a clue of being in presence of partially free rotating molecules. On the other hand, it has been recognised that the free rotational movement of the hydrogens influences the disordered nature of the structure, together with the flexibility in the molecular ring[3,19].

From the analysis of the molecular centres of mass distributions in the predicted structures, we assume that a model with spherical symmetry is not appropriate to describe a freely rotating cyclopentane molecule in the disordered phases. Indeed, some anisotropy in the molecular rotations around the centre of mass and the presence of preferred molecular orientations, may characterise the disorder phases of cyclopentane.

Although CSP was not successful in predicting *a priori* all the experimentally observed crystal structures of cyclopentane, it gave an encouraging result by reproducing the ordered monoclinic crystal structure of phase III. We observed a discrepancy between the calculated (at T = 0K) and the experimental unit cell parameters, which are extrapolated at 0K, which is owing to the parameters[20] of the potential used in the CSP, which were fitted at room temperature. On the other hand, the presence of disorder in the system would make the CSP very difficult even using a more sophisticate intermolecular potential, because of the impossibility of recognising, among such highly dense minima in the lattice energy, those corresponding to the experimental observed structures. It is even possible that the two experimental disordered phases of the cyclopentane correspond to structures, resulting of an average configuration of many predicted structures, which are all very close in lattice energy.

Finally, it is worthy to underline the accuracy of the intermolecular potential in relaxing and predicting the structure of the ordered monoclinic phase III, although it implies a molecular rigid model, which is in contrast with the real flexible nature of the cyclopentane molecule. The good quality of the potential, which has been initially assessed with the use of DMA[11] for the electrostatic contributions, slightly increases with using *ab initio* derived atomic point charges. The latter will be used in a further study of the phase transition in cyclopentane solid state (chapter 7) with molecular dynamics simulations (section 3.5), in which the possibility of running long-time scale simulations (with simulation time up to *ns*) would be severely reduced if we modelled the electrostatics with such as DMA, which has a high computational cost.

6.5 References

(1) Bhatia, P.; Pramanik, P. "CYCLOPENTANE Application:Globalisation, Competitiveness and Flexibility: Case study on Ecofrig project"; Proceedings of the International Conference on Management of Technology, 1997.

(2) Post, B.; Schwartz, R. S.; Frankuchen, I. *J.Phys.Chem.Solids* **1951**, *73*, 5113.

(3) Cardini, G.; Ricci, M.; Righini, R.; Califano, S. *Chem. Phys.* **1994**, *189*, 17.

(4) Torrisi, A.; Leech, C. K.; Shankland, K.; Ibbersen, W. D. R.; Benet-Buchholz, J.; Boese, R.; Leslie, M.; Catlow, C. R. A.; Price, S. L. *(in preparation)* **2007**.

(5) Webster, D. S.; Hoch, M. J. R. *J. Phys. Chem. Solids* **1976**, *37*, 351.

(6) Frisch, M. J. T., G.W.; Schlegel,H.N.; Scuderia,G.E.; Robb,M.A.; Chesseman,J.R.; Zakrzewski,V.G.; Montgomery,J.A.; Stratmann,R.E.; Burant,J.C.; Dapprich,S.; Millam,J.M.; Daniels,A.D.; Kudin,K.N.; Strain,M.C.; Farkas,O.; Tommasi,J.; Barone,V.; Cossi,M.; Cammi,R.; Mennucci,B.; Pomelli,C.; Adamo,C.; Clifford, S.; Ochterski,J.; Petersson,G.A.; Ayalla,P.Y.; Cui,Q.; Morokuma,K.; Malick,D.K.; Rabuck,A.D.; Raghavachari,K; Foresman,J.B.; Cioslowski,J.; Ortiz,J.V.; Stefanov,B.B.; Liu,G.; Liashenko,A.; Piskorz,P.; Komaromi,L.; Gomperts,R.; Martin,R.L.; Fox,D.J.; Keith,T.; Al-Laham,M.A.; Peng,C.Y.; Nanayakkara,A.; Gonzalez,C.; Challacombe,M.; Gill,P.M.W.; Johnson,B.G.; Chen,W.; Wong,M.W.; Andres,J.L.; Head-Gordon,M.; Replogle,E.S.; Pople,J.A.; *GAUSSIAN 98*; Gaussian Inc.: Pittsburgh, 1998.

(7) Moeller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(8) Holden, J. R.; Du, Z. Y.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.

(9) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628.

(10) Williams, D. E.; Cox, S. R. *Acta Crystallogr. Sec. B* **1984**, *40*, 404.

(11) Stone, A. J.; Alderton, M. *Mol.Phys.* **1985**, *56*, 1047.

(12) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(13) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.

(14) Price, S. L. *Adv. Drug Deliv. Rev.* **2004**, *56*, 301.

(15) Ashcroft, N. W.; Mermin, N. D. *SOLID STATE PHYSICS*Orlando, Florida, 1976.

(16) Spek, A. L. *J.Appl.Crystallogr.* **2003**, *36*, 7.

(17) Chisholm, J. A.; Motherwell, S. W. D. *J. Appl. Crystallogr.* **2005**, *38*, 228.

(18) Bruno, J.; Cole, J. C.; Edgington, P. R.; Kessler, M. K.; Macrae, C. F.; McCabe, P.; J., P.; Taylor, R. *Acta Cryst. Sec. B* **2002**, *58*, 389.

(19) Mack, J. W.; Torchia, D. A. *J. Phys. Chem.* **1991**, *95*, 4207.

(20) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

Chapter 7

PHASE TRANSITION IN CYCLOPENTANE SOLID STATE: A CASE OF *ORDER-DISORDER* PHASE TRANSITION

7.1 Introduction

As already introduced in chapter VI, solid cyclopentane is characterised by having three different phases: a monoclinic ordered phase III and two plastic phases (II and I), characterised by the presence of orientational disorder. The structures of those two phases have not been completely solved, especially concerning the intermediate phase II, whose structure is still unknown. The X-ray powder pattern of phase I was, however, partially solved and the structure has a hexagonal unit cell with known cell parameters[1]. A neutron diffraction analysis of all cyclopentane solid phases, which was performed by Dr. C.K. Leech, at the Rutherford Laboratory[2], explores the phase diagram of this compound, by studying the system as function of temperature and pressure, and a further investigation was also performed through synchrotron radiation, recording and X-ray powder patterns[2] over a temperature range between 100K and 144K at P = 1bar.

On the other hand, structural model using static lattice energy minimisation, shows distinguishing features, which can interpreted in the light of the presence the of orientational disorder in the system (chapter 6).

The following chapter will discuss the results obtained from a MD simulation study of cyclopentane solid state, focusing the attention onto the nature of the phase transition between the different phases and onto the orientational disorder and the rotational dynamics of the molecules in the system at different temperatures. We aim to reproduce the main order-disorder phase transition observed in the system and to understand in details the similarities and the differences in the disordered nature of the two plastic phases (I and II) and to give a detailed description of the monoclinic crystalline phase III. Moreover a dynamical study of the system can validate or invalidate certain qualitative information available from the previous crystal structure prediction of the compound,

performed using static lattice energy minimisation within a rigid model of the cyclopentane molecule.

The orientational disorder for plastic crystals has been extensively studied for many compounds such as cubane[3], adantamane[4], cyclohexane[5] and sulfur hexafluoride[6,7] with many different computational techniques, from Monte Carlo[8] to Molecular Dynamics (MD), as previously discussed in chapter 2. A previous study of the plastic phase I of cyclopentane with MD simulations and with a force field, which can model the flexibility of the molecule[9], has found that the internal flexibility of the molecular ring plays a role in the rotational dynamics of the molecule in the disordered hexagonal phase, with a coupling of the internal puckering of the molecular ring and the rotational motion of the molecule. All the three phases were also investigated experimentally through IR[10] and NMR[11] techniques. The latter technique attempted to characterise the internal molecular dynamics of the two plastic phases of cyclopentane, indicating a partial disordered rotational motion of the molecules in the intermediate phase II and a more isotropic rotational disorder in the hexagonal phase I.

In our study of this system with MD simulations at different temperatures, we attempted to use a flexible model for the molecule, but, for reasons that will be illustrated in the next section, we decided to use a rigid model approximation for cyclopentane molecule, which, although does not allow a detailed description of the real rotational dynamics of the molecule in the disordered phase I, as described in the study mentioned above[9], still allows us to give an interesting insight into the nature of the orientational disorder, present in phase II and phase I and to give a better interpretation to the results obtained from the crystal structure prediction in chapter 6.

7.2 Methods

The MD simulations were performed using the DL_POLY_2 program[12], which runs efficiently on parallel architectures. This feature is very important considering the fact that a simulation of crystal structures phase transition requires long time-scale simulations, of the order of 1-8ns. The program solves Newton's equations for a certain system, using the Verlet Leap-Frog algorithm[13,14], previously described in section 3.5.1.

For an accurate study of a phase transition we need to run MD simulations in the anisotropic constant pressure ensemble (NST), which allows the time-evolution of the shape and volume of the unit cell, as previosuly discussed in the MD study of imidazole in section 4.4.2.

We first evaluate the stability of the system at temperature T = 5K, by running 1ns NST simulations and by using two different models for the intermolecular interactions:

1) A simply atom-atom potential (*6-exp* form, reported in section 1.5.2, chapter I), with parameters fitted to organic crystal structures and heats of sublimation (FIT parameters[15]), which implies a rigid body approximation for the cyclopentane molecule.
2) The AMBER force field[16] (reported in section 1.5.2, chapter I), which allows the molecule to maintain its flexibility.

For both the AMBER force field and the *6-exp* potential the electrostatic contribution is given by the atomic *ab-initio* point charges, fitted to the molecular electrostatic potential with the CHelpG algorithm[17](section 3.2).

The calculated atomic charges are reported in table 7.1

Atom Type	*Electronic Charge (e)*
C	0.013880
C	0.018130
C	0.034751
C	0.029270
C	0.003855
H	-0.003979
H	-0.016116
H	-0.022364
H	-0.000309
H	-0.011927
H	-0.007455
H	-0.003656
H	-0.022496
H	-0.017490
H	0.005906

Table 7.1: *Ab initio* calculated atomic point charges for cyclopentane molecule.

Initially, by using the GULP program[18], we relaxed the structure of the monoclinic ordered phase III at 0K, with the cyclopentane molecule in the *ab-initio* optimised

conformation, in order to test the quality of the AMBER force field in modelling this system.

The MD simulations in the NST ensemble were carried out using a Berendsen thermostat-barostat[19] with temperature (τ_T) and pressure (τ_P) relaxation times respectively by τ_T = 0.04 ps and τ_P = 0.01 ps. All the interatomic distances between the atoms in the supercell and the nearest image sites were determined within a cutoff of 10Å. The electrostatics interactions were calculated using the Ewald's summation[20] with a precision of $1 \cdot 10^{-6}$, a cut-off distance $r_{cut\text{-}off}$=10Å for the direct sum, while for the sum in the reciprocal space k_{max} corresponds to the larger vector in the reciprocal space.

The NST simulations using AMBER force field[16]are characterised by the presence of instability in the system. In particular, we observed very strong oscillations in the total energy of the system, although an equilibration time of 200 ps was used in the simulation at 5K and as the simulation evolved in time, a drift in the total and configurational energy of the system was also observed, whose cause remained unclear. Hence, although the AMBER force field showed a relatively good capability in reproducing the monoclinic ordered phase III of cyclopentane at 0K, we did not use it in the following MD study of the phase transition, instead we decided to use the *6-exp* rigid model potential with FIT parameters[15].

In order to calculate at each temperature the properties of the system, such as X-ray powder patterns, mean square displacements (MSD) or orientational distribution functions (ODF), we run simulations of the system in the constant volume and constant energy (NVE) ensemble, in order to avoid fluctuations in the lattice cell parameters, which may affect the accuracy of the calculations. The simulations were performed for 120-150ps, of which 100ps are spent to equilibrate the system. The trajectory file is very dense because of the evaluation of the configuration of the system every 0.1 ps. The starting structure for the phase transition NST simulations is the predicted monoclinic ordered phase I at 0K (*ak2*), which was already discussed in chapter 6. From this structure we created a 3x5x3 supercell, containing 180 molecules, and having the following cell parameters:

	a/Å	*b/Å*	*c/Å*	*α/°*	*β/°*	*γ/°*
***ak2* (353)**	28.218	28.09	30.621	90.00	70.14	90.00

Table 7.2: 3x5x3 supercell of the predicted monoclinic ordered phase III at 0K *(ak2).*

Owing to the limited size (180 molecules) of the simulation box, there are some limitations in the model, which concern the possible influence of the periodic boundary conditions on the system, which have been demonstrated to introduce a residual contribution to the stress in small systems[21] and which can affect the quality of the simulation of the phase transition, as previously discussed in section 3.4. Nevertheless this size of the supercell has been considered a good compromise to permit the system to run for long times (~ *ns*) in a reasonable amount of computer time. The choice of the 3x5x3 supercell also allows us to keep all the three directions of the cell as similar as possible, so that the periodic boundary conditions affect the system isotropically. In all of the simulations we use parallelepiped boundary conditions, which is the most suitable for simulating crystal structures with a general low symmetry unit cell. During the MD study, the pressure of the system has always been kept constant at 1bar

7.3 Results

7.3.1 Attempts to use a molecular flexible model for the study of the phase transition.

As already introduced in the previous section, we initially performed a relaxation of the predicted structure of the monoclinic phase III of cyclopentane, with the molecule in the optimised geometry, using the AMBER force field[16] and the GULP program[18].
The results of the relaxation are reported in table 7.3.

	a/Å (*δ*%)	*b/Å* (*δ*%)	*c/Å* (*δ*%)	*β/°*	*Vol/Z* */Åmol⁻¹*	*Lat.En.* */kJmol⁻¹*
ExpOpt	9.428	5.429	10.145	111.63	120.68	-43.530
ExMiOpt	8.989 (-4.66)	5.227 (-3.72)	9.753 (-3.39)	110.46	107.34	-47.114

Table 7.3: Structural data for the relaxation of ordered monoclinic phase III of cyclopentane at 0K, by using AMBER force field. *ExpOpt* indicates the experimental structure from single crystal diffraction, crystallised by prof. R.Boese, with the molecules in the ab initio optimised conformations. *ExMiOpt* indicates the same structure after lattice energy minimisation. The root mean square coefficient (*RMS%*) were calculated as: $RMS\ (\%) = \sqrt{\frac{(\delta a)^2 + (\delta b)^2 + (\delta c)^2}{3}}$, where δa (and similarly for δb and δc) has the expression: $\delta a = 100 \cdot \frac{(a_{exper.} - a_{calc.})}{a_{calc.}}$.

An RMS value by ~ 3.95%, which represents an acceptable deviation of the lattice parameters from experiment, indicates that the force field performs a reasonably well lattice energy minimisation of the cyclopentane experimental monoclinic structure. In evaluating the lattice energy the AMBER force field[16] gives a considerably lower result (a discrepancy $\Delta U_{\text{AMBER-FIT}}$ ~ 7.5 kJ mol^{-1}) than the value obtained by energy minimisation with the *6-exp* potential using FIT parameters[15] (see table 6.4, chapter 6). This is due to the treatment of the molecule as flexible by the AMBER force field[16], which hence, also takes in account also the intra-molecular contribution, which is appreciable and not considered by the rigid body approximation implied by the use of the *6-exp* potential with FIT parameters[15].

In order to examine the variation of the fluctuations in temperature and pressure during the MD simulation, we performed two MD simulations on the optimised predicted crystalline system at 5K in NST ensemble with AMBER force field, using a Berendsen thermostat and barostat[19]: the first simulation with relaxation times τ_T = 1.0 ps and τ_B = 0.1ps and the second with lower values: τ_T = 0.04ps and τ_B = 0.01ps. In both cases, we used a timestep of 10^{-4} ps, to make the system stable, particularly with respect to the C-H stretching motion in the cyclopentane molecule[22,23](11.3 fs $\leq \nu_{stretch} \leq$11.7 fs). The total time of the simulations is 1ns, divided into 100ps for the equilibration with scaling of the velocities every 0.5 ps, and 900 ps for the time-evolution of the system. In figure 7.1, 7.2, 7.3 and 7.4 we show the total internal energy of the system and the fluctuations of the temperature as function of time respectively in the two simulations.

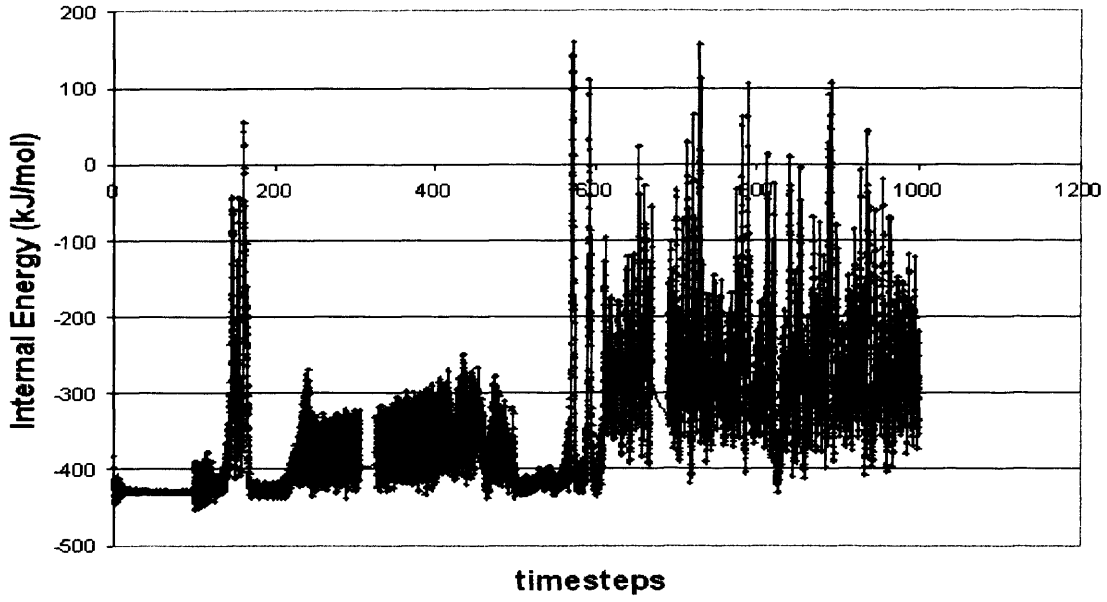


Fig. 7.1: Internal energy *vs* time in the MD simulated phase III of cyclopentane at 5K, using AMBER Force Field and Berendsen thermostat with τ_T = 1.0ps and barostat with τ_B = 0.1ps.

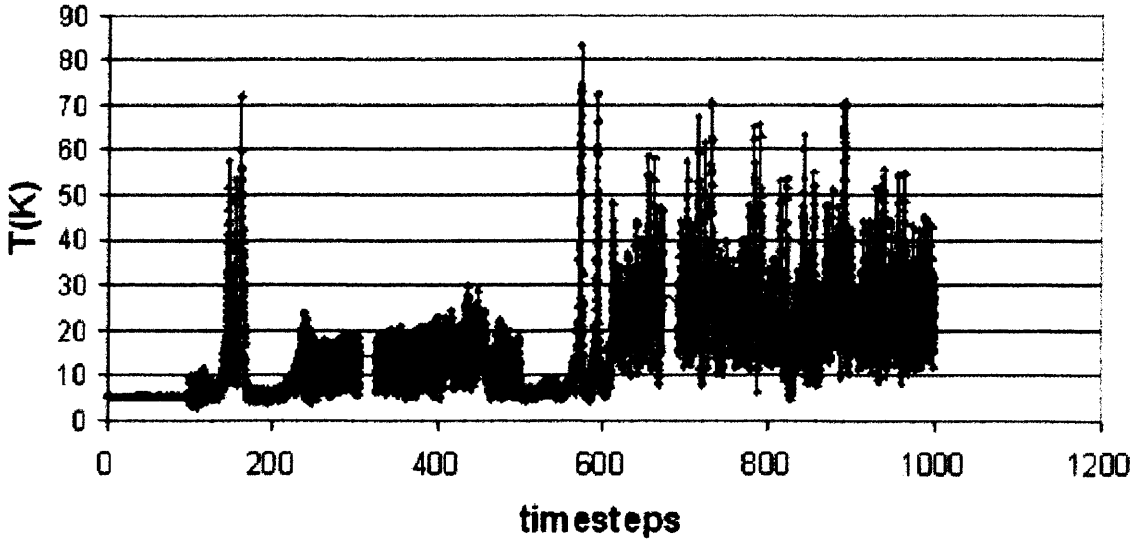


Fig. 7.2: Temperature *vs* time in the MD simulated phase III of cyclopentane at 5K, using AMBER Force Field and Berendsen thermostat with τ_T = 1.0ps and barostat with τ_B = 0.1ps.

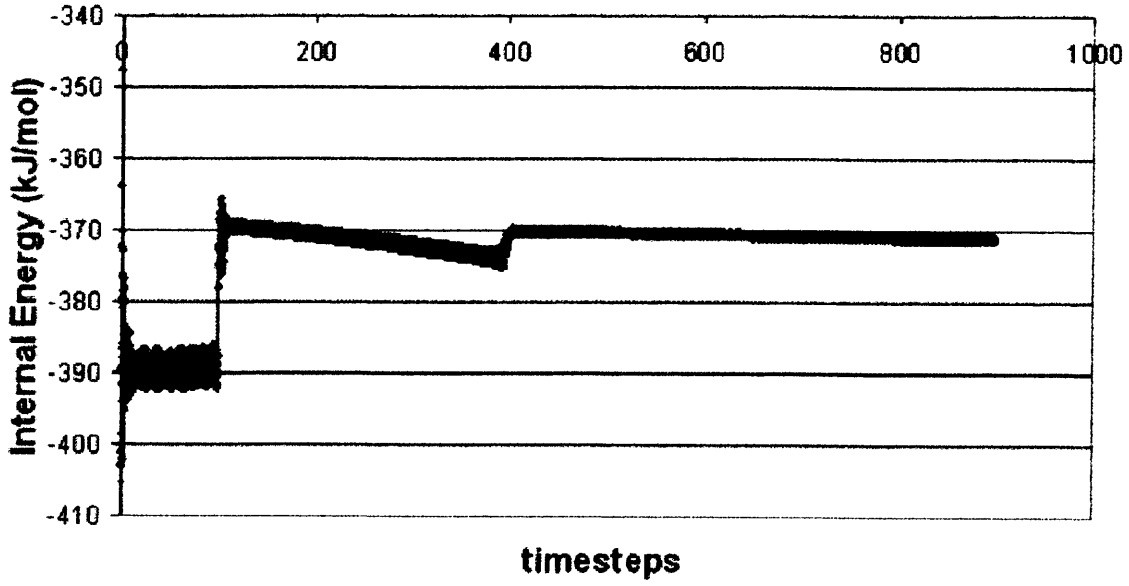


Fig. 7.3: Internal energy *vs* time in the MD simulated phase III of cyclopentane at 5K, using AMBER Force Field and Berendsen thermostat with τ_T = 0.04ps and barostat with τ_B = 0.01ps.

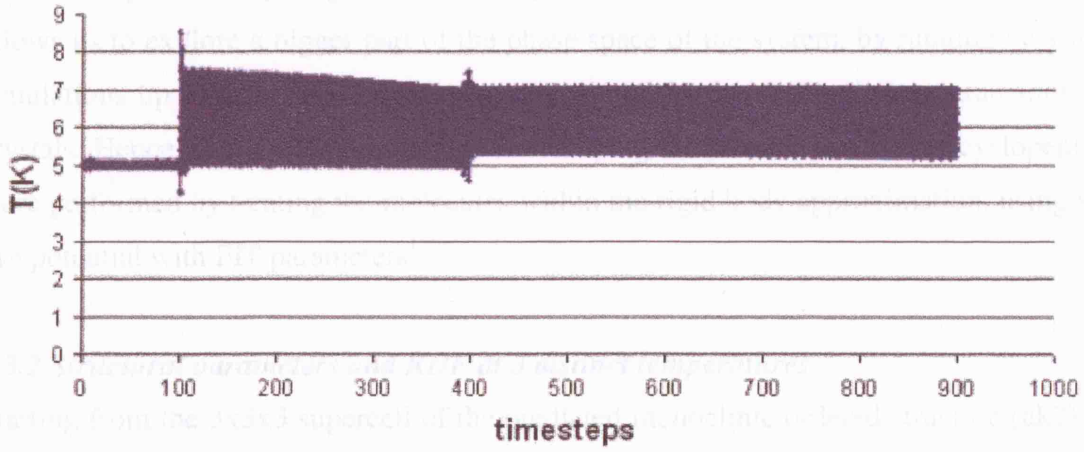


Fig. 7.4: Temperature *vs* time in the MD simulated phase III of cyclopentane at 5K, using AMBER Force Field and Berendsen thermostat with τ_T = 0.04ps and barostat with τ_B = 0.01ps.

In the simulation with a Berendsen thermostat with longer relaxation times (τ_T = 1.0ps and τ_B = 0.1ps), it clearly appears that the total internal energy of the system suffers very large fluctuations (figure 7.1), due to the presence of strong fluctuations in the temperature (figure 7.2), which reflect a strong instability of the system. We obtained a bigger improvement in the stability of the system simulated with a flexible molecular model, using shorter relaxation times for the Berendsen thermostat-barostat (τ_T = 0.04ps and τ_B = 0.01ps). In this case the total internal energy (figure 7.3) suffers much smaller fluctuations (together with temperature, reported in figure 7.4), and it remains in the range of negative values. Nevertheless, we still observe unstable behaviour of the energy, which increases by 20 kJ/mol after the equilibration time and shows sudden jumps after long periods of time. The reason of such behaviour of the MD simulated system within the AMBER force field at 5K is not clear and we think that the instabilities might imply the necessity of a longer equilibration time for the system, which would make our simulations more time-consuming, because of the use of a very small time step (10^{-4}ps).
On the other hand, we found good stability of all the thermodynamics quantities, by simulating the system with a rigid molecular model, using a *6-exp* potential with room temperature fitted parameters (FIT[15] potential), and using a Berendsen thermostat[19] with small values of relaxation times (τ_T = 0.04ps and τ_B = 0.01ps). With the latter potential, at any temperature, the stability of the system is fully reached in an equilibration time of 50-100ps and the possible choice of a timestep of 0.001ps (as we do not need to consider the C-H vibrational intramolecular motion[22] within a rigid model) allows us to reduce

considerably the computing time necessary for a MD simulation of the system, which allows us to explore a bigger part of the phase space of the system, by running very long simulations up to 4ns. This requirement is essential in the study of phase transitions of crystals. Hence all the MD simulations for studying the phase transition of cyclopentane were performed by treating the molecules within the rigid body approximation, using a *6-exp* potential with FIT parameters[15].

7.3.2 Structural parameters and RDF at 3 distinct temperatures.

Starting from the 3x5x3 supercell of the predicted monoclinic ordered structure (ak2) we initially performed three different simulations in the NST ensemble for 3ns with an equilibration time of 100ps respectively at three different temperatures:

- 100K: temperature at which the ordered monoclinic phase III persists.
- 150K and 170K: temperatures at which the orientationally disordered hexagonal phase I should be stable.

The structural changes with temperature in the supercell are reported in table 7.4:

T(K)	*a*/Å	*b*/Å	*c*/Å	α/°	β/°	γ/°	Vol/Z/Å^3
0 *(ak2)*	9.406	5.618	10.207	90.00	70.14	90.00	126.86
100	9.134	5.762	10.307	90.00	75.22	90.00	131.13
150	9.511	5.917	10.276	89.99	89.66	90.02	144.19
170	9.577	5.952	10.326	89.99	90.03	90.06	147.10

Table 7.4: Structural cell parameters of cyclopentane from the MD simulated 3x5x3 supercell at 100K 150K and 170K. In the first row we report the structural parameters of the starting structure. *ak2* is the predicted crystals structure (at 0K), corresponding to the experimental monoclinic phase III.

The most important change in the lattice parameters concerns the β cell angle, which changes from 70.14° to 90.03° and is accompanied by a small change in all of the three cell vectors, by values comprised in the range between 0.1 and 0.4 Å. These changes lead to a transformation of the unit cell of the system from a monoclinic (observable at 0K and 100K) to an orthorhombic unit cell (at 170K). We note that is possible to define an equivalent orthorhombic cell for the experimental hexagonal unit cell of cyclopentane phase I[24], as schematically illustrated in figure 7.5.

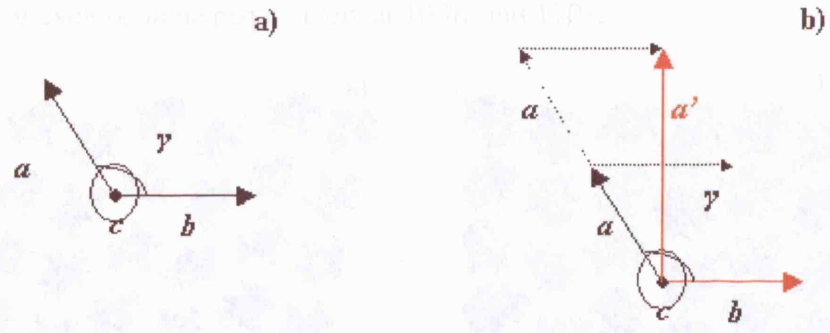


Fig. 7.5:a) Conventional hexagonal unit cell **b)** equivalent orthorhombic cell

Table 7.5 reports the values of the lattice parameters of the two equivalent cells of the cyclopentane phase I for the experimental structure and that simulated at 150K.

	a/Å	*b*/Å	*c*/Å	α/°	β/°	γ/°
Hegonal Exp	5.830	5.830	9.330	90	90	120
Equivalent Orthorombic Exp	*a'*/Å	*b*/Å	*c*/Å	α/°	β/°	γ/°
	10.10	5.830	9.330	90	90	90
Orthorombic MD simulated 150K	9.511	5.917	10.276	89.99	89.66	90.02

Table 7.5: Experimental cell parameters of the hexagonal cyclopentane phase I and its equivalent orthorhombic cell.

If we swap *a* and *c* axes in the MD simulated system we can observe (in table 7.5) that the cell parameters of the MD simulated structure at 150K are in good agreement with the experimental parameters of phase I in the equivalent orthorhombic form. Therefore, as regards cell dimensions, the MD simulation can reproduce the high-T disordered phase I, starting from the ordered monoclinic phase III to a good level of approximation. The disordered nature of the phase I is also modelled by the MD simulation. In figure 7.6 we

can clearly observe the different symmetry of the two MD simulated phases, phase III and phases I of cyclopentane respectively at 100K and 170K.

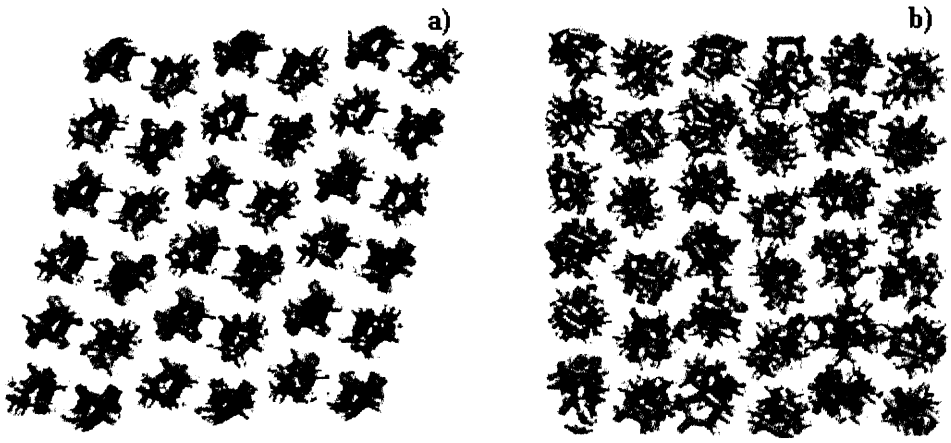


Fig. 7.6: Snapshots of the two phases of cyclopentane from the MD simulation: **a)** monoclinic ordered phase III at 100K.**b)** hexagonal (equivalent orthorhombic) disordered phase I at 150K. In both of the cases the view is along the *c* axis.

We can see the presence of disorder in the high temperature phase I by examining the superposition of the molecules along the *c* axis in the supercell of the system shown in figure 7.6b, in which the molecules appear in freely rotated positions around their centres of mass, and have many different orientations. This configuration is different compared to that observable by the superposition of the molecules in the system at 100K (figure 4a), with all the molecules having the centres of mass along the same line, parallel to the *c* axis, and with the same orientation.

The analysis of the Radial Distribution Functions (RDFs), introduced in chapter 4, can be very helpful in distinguishing an ordered crystalline system, from a disordered liquid-like one. We plotted the values of the RDF relative to the couple of atoms, C-C, C-H and H-H in the system at the three different temperatures, as illustrated respectively in figures 7.7, 7.8 and 7.9.

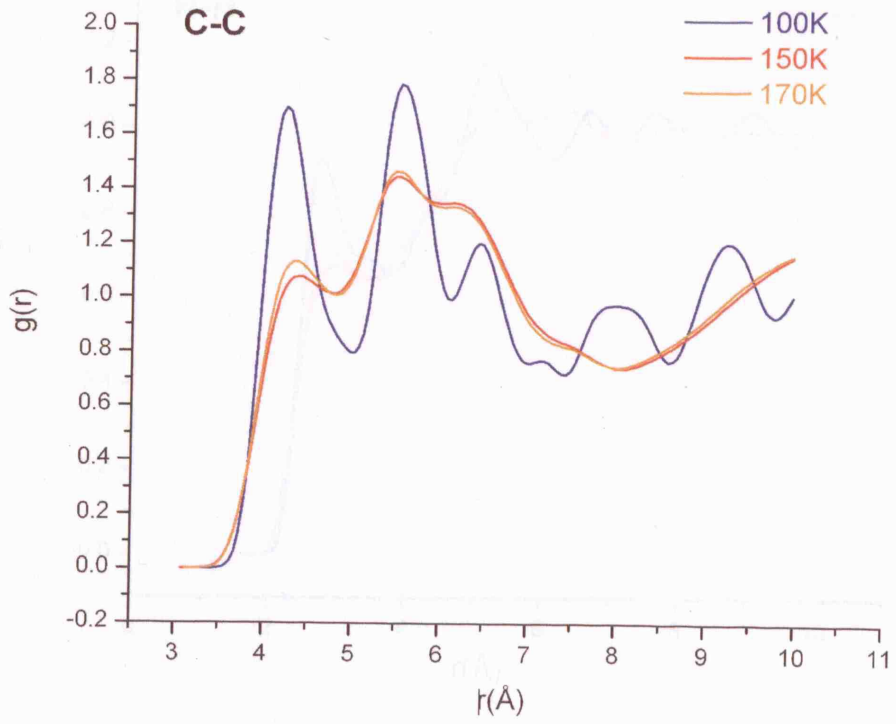


Fig. 7.7: C-C RDF of MD simulated cyclopentane (3 *ns*) at 3 different temperatures.

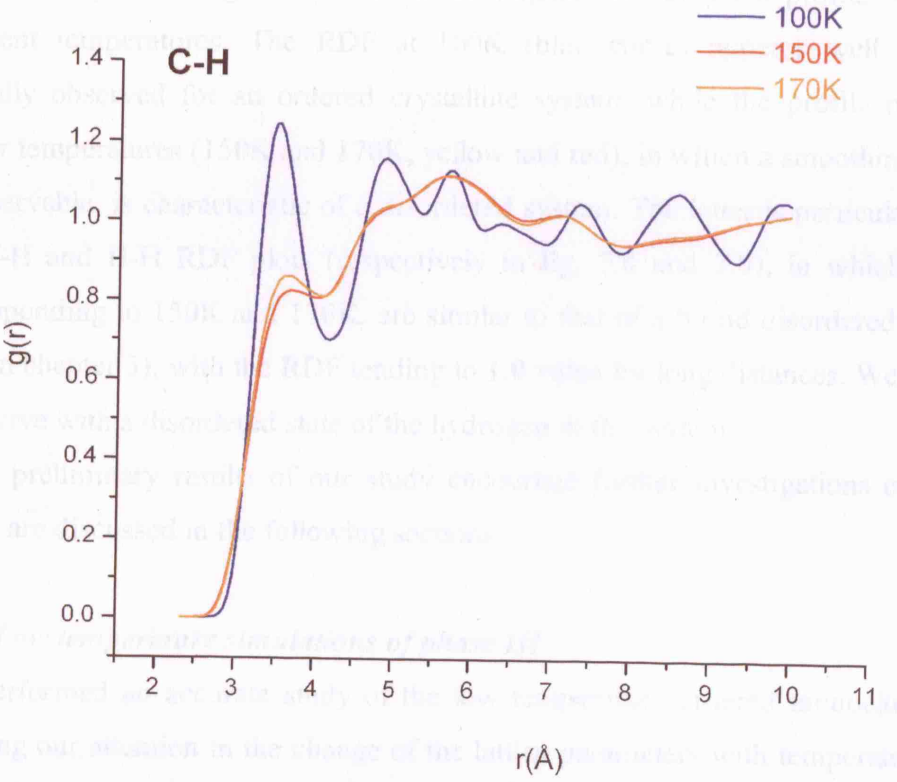


Fig. 7.8: C-H RDF of MD simulated cyclopentane (3 *ns*) at 3 different temperatures.

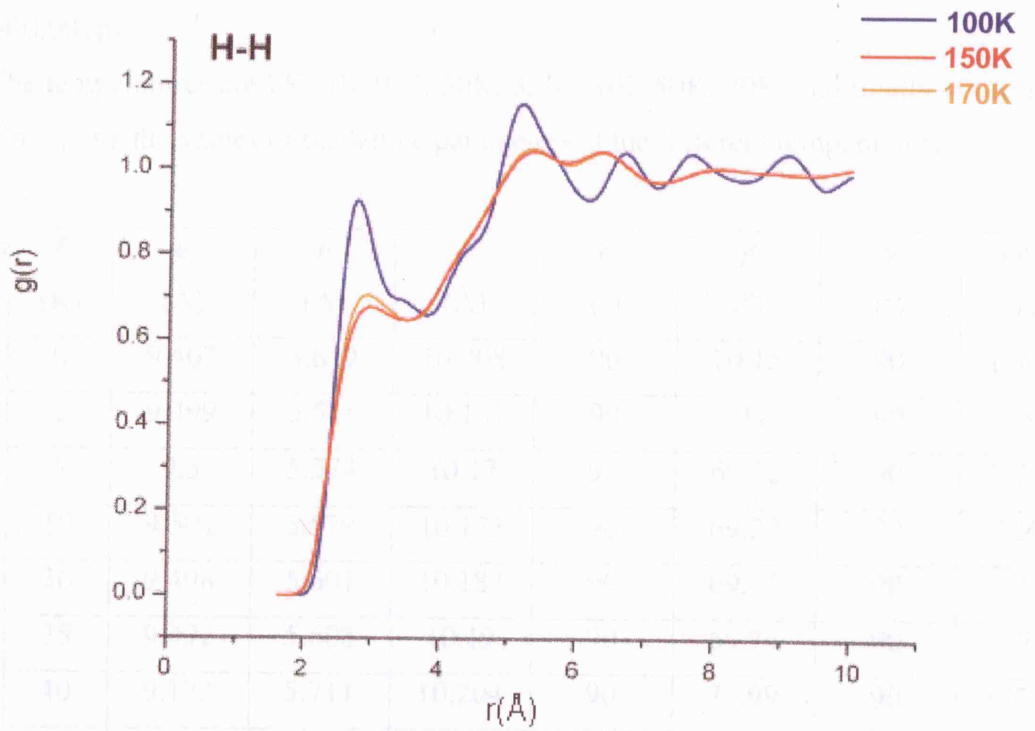


Fig. 7.9: H-H RDF of MD simulated cyclopentane (3 *ns*) at 3 different temperatures.

In the RDF plots in figure 7.7 we can distinguish two different profiles of RDF at the different temperatures. The RDF at 100K (blue curve) appears well structured, as typically observed for an ordered crystalline system, while the profile of the RDF at higher temperatures (150K and 170K, yellow and red), in which a smoothing of the peaks is observable, is characteristic of a disordered system. The latter is particularly evident in the C-H and H-H RDF plots (respectively in fig. 7.8 and 7.9), in which the profiles, corresponding to 150K and 170K, are similar to that of a liquid disordered phase (figure 3.11 in chapter 3), with the RDF tending to 1.0 value for long distances. We can associate this curve with a disordered state of the hydrogen in the system.

These preliminary results of our study encourage further investigations of the system, which are discussed in the following sections.

7.3.3 Low temperature simulations of phase III

We performed an accurate study of the low temperature ordered monoclinic phase III, focusing our attention in the change of the lattice parameters with temperature. The NST simulations are performed using a Berendsen[19] thermostat and barostat with respective relaxation times τ_T = 0.04ps and τ_B = 0.01ps. The simulations run for a total time of 3ns,

of which 100ps are equilibration time and 2900ps the production time, using a timestep of 0.001ps.

The temperatures are 2K, 5K 10K, 30K, 35K, 40K, 50K, 70K, and finally 100K and table 7.6 reports the values of the lattice parameters at the different temperatures.

T (K)	*a* (Å)	*b* (Å)	*c* (Å)	*α* (°)	*β* (°)	*γ* (°)	*Vol/Z* (Å^3)
0	9.407	5.619	10.208	90	70.15	90	126.86
2	9.499	5.571	10.167	90	69.8	90	125.77
5	9.5	5.574	10.17	90	69.22	90	125.89
10	9.501	5.579	10.173	90	69.29	90	126.1
30	9.498	5.601	10.187	90	69.61	90	126.99
35	9.492	5.608	10.19	90	69.74	90	127.22
40	9.122	5.711	10.204	90	73.99	90	127.74
50	9.144	5.721	10.22	90	74.3	90	128.25
70	9.122	5.735	10.253	90	74.64	90	129.31
100	9.134	5.762	10.307	90	75.22	90	131.13

Table 7.6: Lattice parameters *vs* temperature for the low temperature monoclinic ordered phase III of cyclopentane.

Examining the data in table 7.6 relating to the cell vectors *a*, *b* and *c* and *β* angle, we note an interesting behaviour: while all the cell parameters remain approximately constant between 2K and 35K, with *β* angle changing at most by 0.5°, a sudden change occurs at 40K for *a*, *b* and *β* cell parameters, with *a* decreasing by 0.4 Å, *b* increasing by up to 0.25 Å and *β* changing by 4°-6°, while *c* changes at most by 0.13 Å, following a very small expansion. Figure 7.10 shows the change of the relative parameters defined by the following formula:

$$relative\ cell\ par. = \frac{cell\ par.(T)}{cell\ par.(2K)} \quad . \qquad [7.1]$$

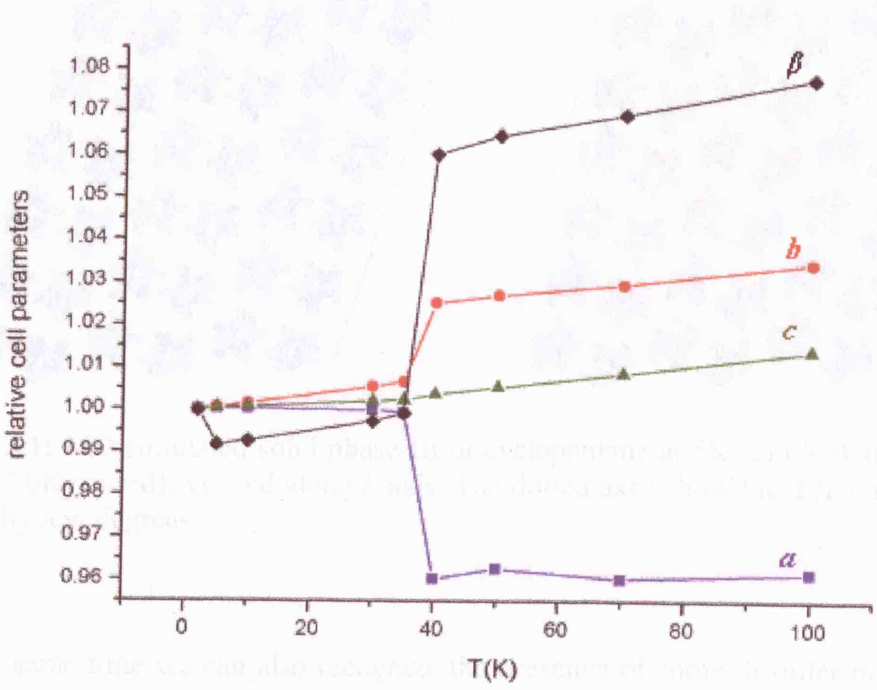


Fig. 7.10: Relative cell parameters *vs* temperature from MD simulations of a 3x5x3 supercell of ordered monoclinic phase III of cyclopentane: *a, b, c* axes and β angle.

All the three cell vectors and the β-angle are approximately constant from 2K to 3K, with small changes due to the effect of the temperature, but in a very small range of temperatures (5K), *b* and β increase suddenly and *a* decreases rapidly, with *c* cell vector being the sole parameter, which undergoes a continuous moderate expansion The structural behaviour, shown in figure 7.10, clearly indicates a phase transition between 35K and 40K, which is followed by a thermal expansion of the crystal.

This small phase transition consists of a distortion of the monoclinic as shown in figure 7.11, which indicates a quite appreciable change in the monoclinic unit cell at the higher temperature (in red), with β opening by ~ 5-6°, which is the key parameter in the distortion of the unit cell.

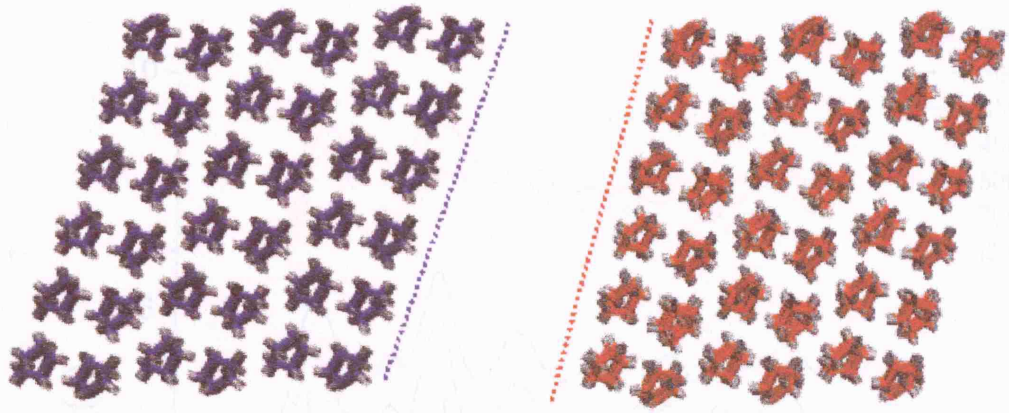


Fig. 7.11: MD simulated solid phase III of cyclopentane at 5K (in blue) and the distorted cell at 70K (in red), viewed along *b* axis. The dotted axes show the difference in the β cell angle by few degrees.

At the same time we can also recognize the presence of some disorder in the orientation of the molecules in phase III at the higher temperature (red structure), which are not perfectly iso-oriented as in the low temperature range between 2K and 35K. This effect is due to the higher librational motion of the molecules around their centres of mass, at higher temperatures.

The presence of a subtle phase transition is also observable in the C-C pair distribution function (RDF) of the system shown in figure 7.12, the main peak positions of which are reported in table 7.7.

T (*K*)	*r* (Å)					
2K-35K	4.03	4.93	5.48	6.38	8.69	9.46
40K-100K	4.22	No peak	5.57	6.44	No peak	9.25

Table 7.7: Peak positions in the C-C radial distribution function of the solid phase III of cyclopentane, in two different intervals of temperatures: 4K-35K and 40K-100K.

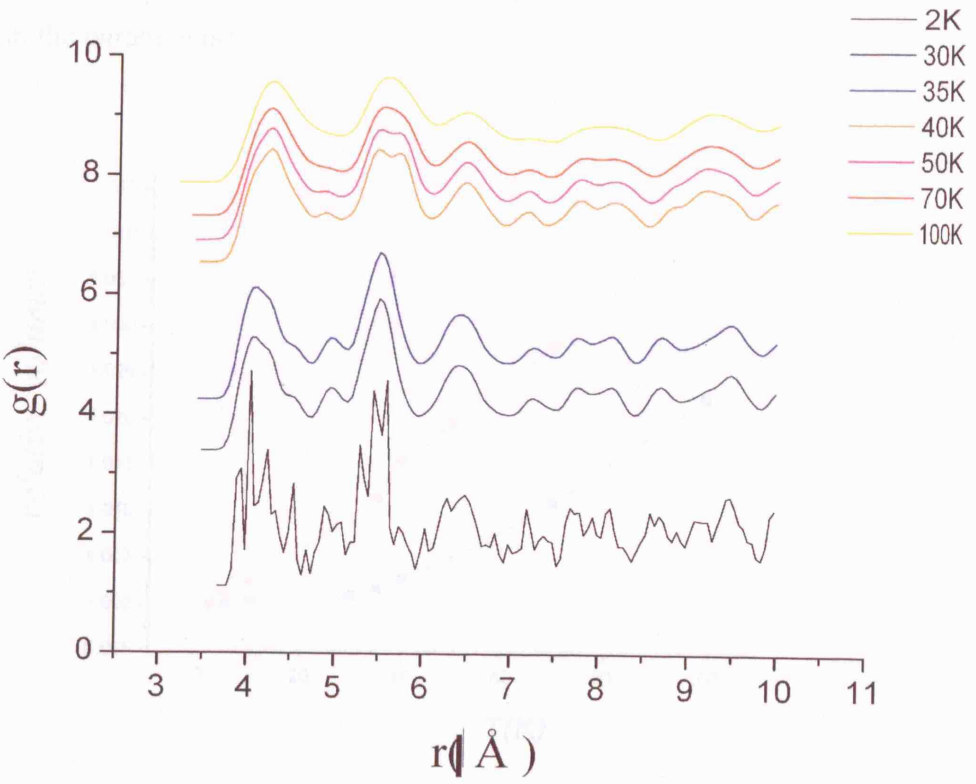


Fig. 7.12: Evolution of the C-C RDF with temperature, in the interval 2K-100K for cyclopentane ordered phase III.

The change in the solid phase III unit cell between 35K and 40K is reflected in the very small changes of some peaks in the RDF. From figure 7.12, we note the on-set of slight molecular motion (libration) activated by temperature, starting from 30K. Nevertheless the profile and the position of the peaks of the starting RDF at 2K are maintained up to 35K, while some changes appear at higher temperatures (40K-100K), which essentially involve a shift of some peaks starting from 40K with respect to the peaks at 2K. In table 7.7 is shown that the RDF peaks at 4.93Å and at 8.69 Å observed in the range 2K-35K, disappear in the higher temperatures interval (T > 40K). These subtle changes indicate a slight modification of the structure of the system on raising the temperature from 35K to 40K.

There is no indication of this phase transition in the experiments, which show cell parameters, which remain almost constant up to 35K and change continuously by small percentages, giving as result, a moderate linear thermal expansion of the system, starting only at 40K (figure.7.13 red plot). Hence the observed distortion of the monoclinic cell in the MD simulations might be an artefact of the model and it s limitations, already

mentioned above (PBC and intermolecular potential, which contains intrinsic thermal effect in the parameters).

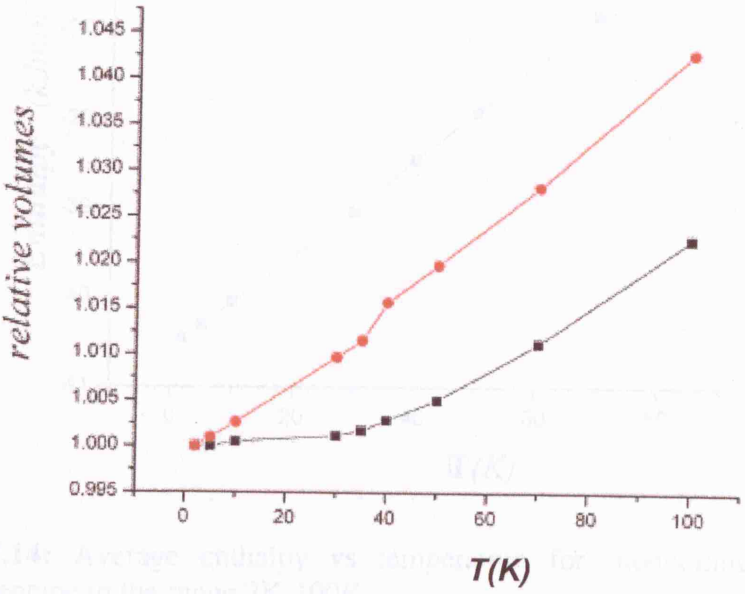


Fig.7.13: Plots of the relative volumes for the calculated (red curve) and experimental (black curve) ordered monoclinic phase III of cyclopentane as functions of temperature. The relative volumes are expressed as $V_i(T) / V_i(2K)$.

On the contrary in the simulated system, we observe a slight discontinuity in the trend of the volume of the MD simulated system with temperature (figure 7.13 red plot) occurring between 35K and 40K. At the same time a strong thermal expansion characterises both intervals of temperatures, 2K-35K and 40K-100K. The thermal expansion is much stronger than that observed in the experimental structure, especially in the system at very low temperature (in the interval 2K-35K) and the cause of this effect may be the interatomic potential, which contains parameters fitted at room temperature. The observed phase transition is subtle and does not involve a big change in the enthalpy of the system (figure 7.14), even if a peak in the specific heat at constant pressure (C_p) is still noted (figure 7.15), indicating the presence of a structural change in the system.

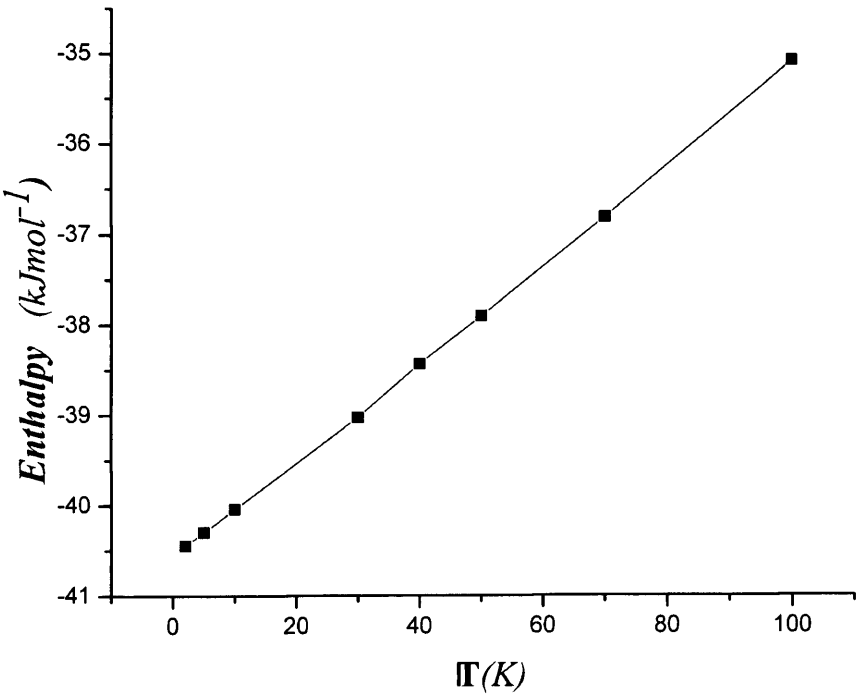


Fig. 7.14: Average enthalpy vs temperature for monoclinic ordered phase III of cyclopentane in the range 2K-100K.

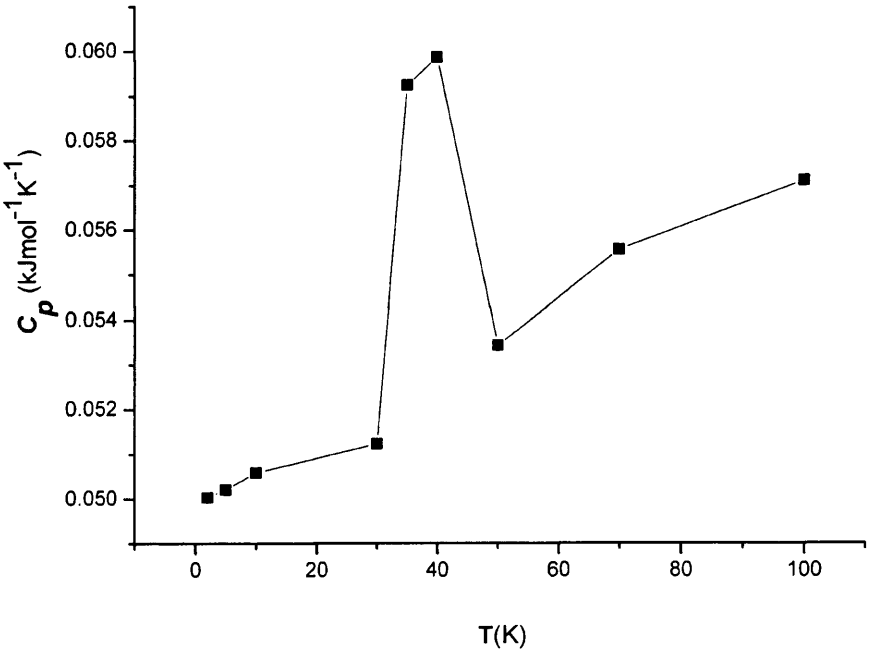


Fig. 7.15: Average specific heat at constant pressure (C_p) *vs* temperature for monoclinic ordered phase III of cyclopentane in the range 2K-100K.

A very approximated estimate of the enthalpy of such transformation (ΔH_t) can be obtained by the integrated area under the peak occurring in the plot of the specific heat as function of temperature (figure 7.15), following the formula[3]:

$$\Delta H_t = \int_{T_1}^{T_2} Cp\, dT \quad . \qquad [7.2]$$

The calculated average enthalpy, calculated by integrating the C_p curve between 30K and 50K, is ΔH_t ~ 1.14kJ mol^{-1}, which is a such small value that we can interpret these structural changes as an effective subtle second order phase transition, not observed in the experiments, which still preserves the symmetry of the system, but the confidence on this value is limited by the very few number of points gathered in the plots (figure 7.15).

7.3.4 Simulated heating of cyclopentane system from 100K to 270K

For a better understanding of the main phase transition between phase III and phase I, through an intermediate disordered phase II, as observed either in the experimental X-ray powder patterns at 118K (phase III-phase II) and 138K (phase II-phase I), or in the preliminary MD calculations at 170K, we performed a careful simulation at several temperatures between 100K and 170K, for 1*ns* (at 110K, 120K, 150K, 160K) and 3*ns* (in the range between 120K and 140K) extending the equilibration time from 50ps up to 100ps. The latter interval is that, in which the presence of the intermediate phase II is likely to be observed.

Table 7.8 reports the values of the structural parameters obtained from the MD simulations at the different temperatures, and their relative trend with the temperature is reported in figure 7.16 and 7.17.

T /K	*a*/Å	*b*/Å	*c*/Å	α/°	β/°	γ/°	Vol/Z/ Å^3	Time/ ns
0	9.406	5.618	10.207	90.00	70.14	90.00	126.86	
100	9.134	5.762	10.307	90.00	75.22	90.00	131.13	1
110	9.132	5.779	10.330	90.17	75.97	89.44	131.96	1
120	9.132	5.796	10.351	90.06	75.81	89.96	132.78	1
122	9.126	5.811	10.353	90.00	75.93	89.95	133.14	3
124	9.104	5.857	10.368	90.00	76.39	90.00	134.31	3
125	9.445	5.875	10.182	89.91	88.52	89.96	141.00	3
126	9.431	5.872	10.205	90.13	88.19	89.19	140.98	3
127	9.493	5.874	10.174	89.98	89.95	90.05	141.78	3
128	9.484	5.879	10.185	89.74	89.98	89.90	141.92	3
130	9.485	5.880	10.199	90.04	90.04	89.93	142.16	3
132	9.506	5.881	10.193	90.12	89.94	90.11	142.41	3
134	9.494	5.890	10.208	90.13	89.90	89.98	142.68	3
136	9.506	5.888	10.216	89.99	89.98	90.10	142.92	3
138	9.507	5.894	10.225	90.03	90.09	90.06	143.18	3
140	9.503	5.900	10.237	90.07	90.03	90.02	143.44	3
150	9.524	5.919	10.267	90.10	90.09	90.02	144.68	1
160	9.557	5.936	10.289	89.97	90.04	90.04	145.89	1
170	9.577	5.952	10.326	89.99	90.03	89.99	147.10	1
200	9.663	6.000	10.406	89.98	89.95	89.91	150.80	1
230	9.747	6.045	10.494	90.02	89.97	89.96	154.53	1
260	9.844	6.098	10.576	90.05	90.05	90.01	158.67	1
265	9.869	6.103	10.603	90.01	90.00	89.93	159.59	1
268	9.903	6.100	10.612	90.02	90.82	90.04	160.12	1
270	9.124	5.892	13.375	74.45	89.66	85.54	169.22	1

Table 7.8: Structural parameters of the MD simulated cyclopentane structures at different temperatures in the range 100K-270K.

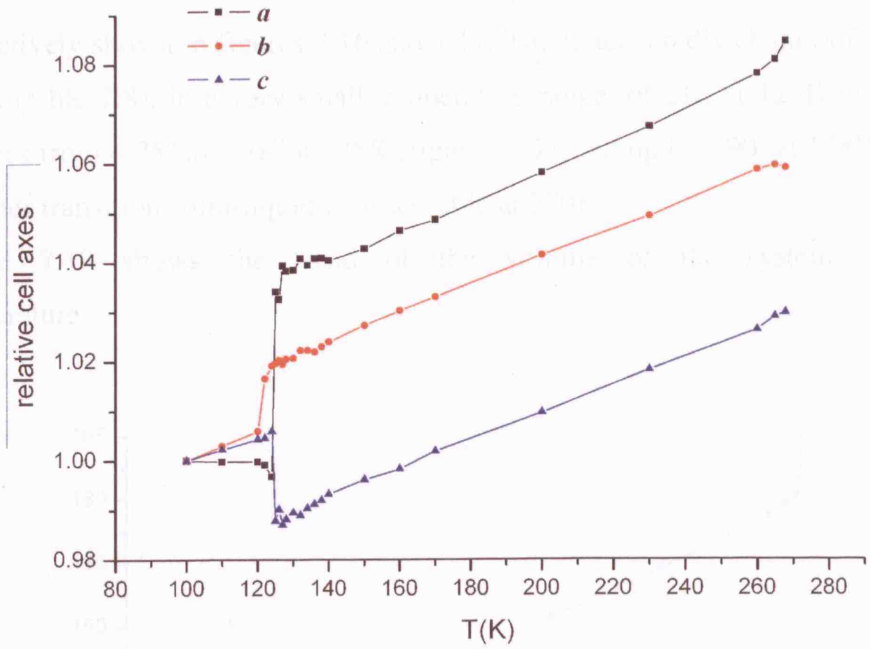


Fig. 7.16: Relatives cell axes *vs* temperature for MD simulated 3x5x3 supercell of the cyclopentane structure in the range 100K-270K. The generic *x* relative cell parameter (where *x* = *a, b, c*) are calculated as it follows: *x*(T) / *x*(100K).

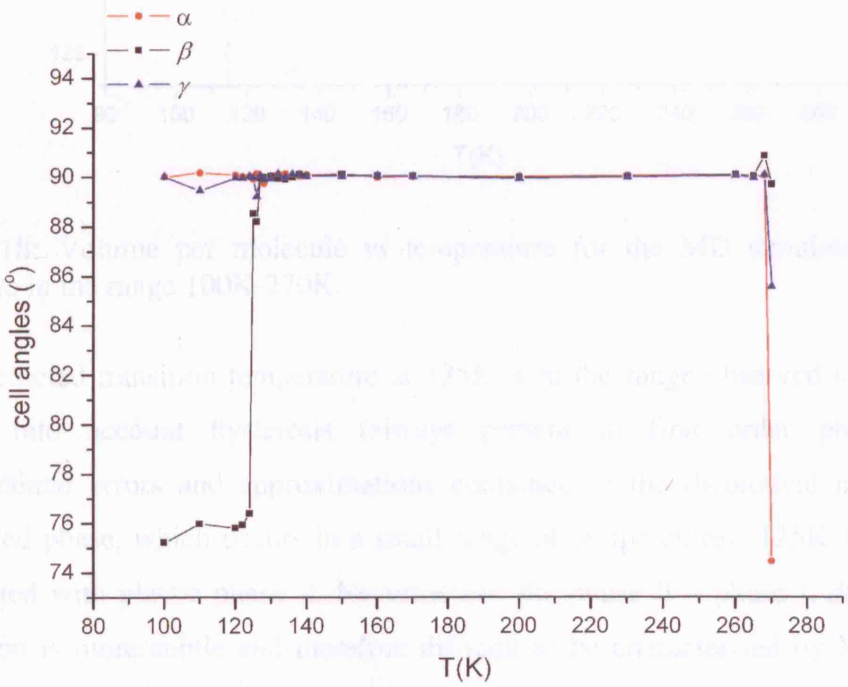


Fig. 7.17: Cell angles *vs* temperature for MD simulated 3x5x3 supercell of the cyclopentane structure in the range 100K-270K.

The key structural parameters, indicating the phase transition from the ordered monoclinic phase III to a disordered new phase, are the β angle and a cell vector, respectively shown in figures 7.16 and 7.17. The latter rapidly changes from ~ 9.1 Å to ~ 9.4 Å (table 7.8), in a very small temperature range, of 2K, at 125K, while the β angle changes from ~ 75° to ~ 88° at 125K (figure 7.17), and up to ~ 90° at 128K.

Another transition solid-liquid is observable at 270K.

Figure 7.18 shows the trend of the volume of the system as function of temperature.

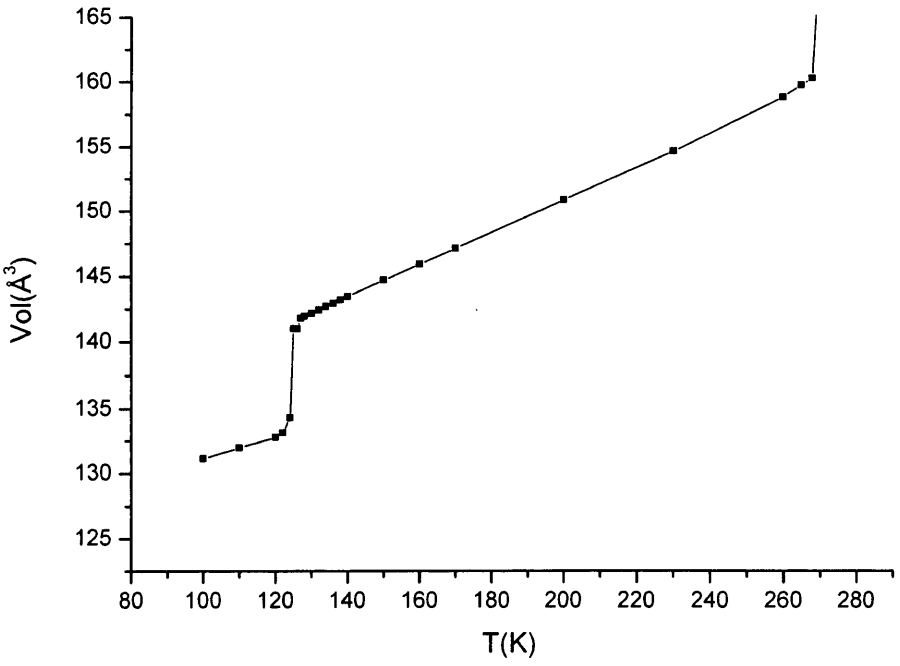


Fig. 7.18: Volume per molecule *vs* temperature for the MD simulated cyclopentane structure in the range 100K-270K.

The predicted transition temperature at 125K is in the range observed in experiments[25], taking into account hysteresis (always present in first order phase transition), experimental errors and approximations contained in the theoretical model. The new simulated phase, which occurs in a small range of temperatures (125K-126K), might be associated with plastic phase II. Nevertheless, the phase II – phase I, *disorder-disorder* transition is more subtle and therefore difficult to be characterised by MD simulations, especially in view of the lack of experimental values for the unit cell of the intermediate phase II, which hinders a clear interpretation of the simulated structural parameters in the intermediate interval of temperatures.

From the calculations, the hexagonal plastic phase I, formally appears already at 130K, with the cell parameters reproducing those of the experimental orthorhombic, equivalent to the hexagonal unit cell. In practice the snapshots of the time evolution of the MD simulated system indicates 140K at which the orthogonal unit cell is definitely stable. Unfortunately the plot of the change in the volume of the system with temperature (figure 7.18), because of the presence of a continuous, linear thermal expansion after the phase transition at 125K-126K, which could mask a further small change due to subtle second order phase II - phase I transition. Nevertheless the plot in figure 7.8 shows a clear *order-disorder*, which is associated with a discontinuity in the volume of the unit cell at 124K-125K, which increases by 6.3% from the ordered monoclinic phase III to the disordered hexagonal phase I.

For both of these two phases we can observe a linear thermal expansion, in a very extended range of temperatures, up to 260K. We calculated the thermal expansion coefficient for the two phases from the slopes $\left(\frac{\partial V}{\partial T}\right)$ of the two different trends of the volume, observed in the plot of the volume with temperature at constant pressure: a linear trend, between 100K and 122K, corresponds to the monoclinic phase III and another steeper linear trend, between 128K and 170K, is associated to hexagonal plastic phase I. The fit of the volumes in these two separated intervals of temperature gives the thermal coefficients α_{therm} for the two phases (table 7.9), calculated according to the definition:

$$\alpha_{therm} = \frac{1}{V} \cdot \left(\frac{\partial V}{\partial T}\right)_P \qquad [7.3]$$

in which V is the volume of the unit cell.

	$\left(\frac{\partial V}{\partial T}\right)_P$ /Å^{3}K^{-1}	*constant* /Å^3	α_{therm} /10^{-6}K^{-1}
monoclinic ordered Phase III	0.0825	122.8	671.8
hexagonal plastic Phase I	0.1254	125.8	996.8

Table 7.9: Thermal expansion coefficient α_{therm} and parameters from the linear fit of the curve Vol/Z vs T, for phase III and phase I of cyclopentane.

As we can see from the data reported in table 7.9, the thermal expansion of the plastic phase I is larger than that of the ordered phase III. This characteristic is quite common in crystals exhibiting a low temperature ordered phase and a higher temperature plastic phase[3]. The larger value for the plastic phase I can be explained by considering the large-amplitude molecular reorientational motion in this phase, which reduces the dispersive interactions between the molecules, allowing an easier thermal expansion, as observed in other disordered crystalline systems[3].

It is interesting to note that the hexagonal plastic phase I of cyclopentane results from the MD simulations being stable up to 265K, with the simulations showing a melting point at 270K, which is far from the experimental value of ~180K. The reason for this disagreement is the presence of super-heating effects in the MD simulated system in bulk. Indeed, experimental melting points are usually observed in solid-liquid phase transitions occurring via heterogeneous nucleation sites at the surfaces of the solid [26,27,28], through the formation of a solid-liquid interface, which involves a free energy barrier, which is difficult to overcome in direct MD simulations[29]. In a simulated system like a bulk crystal, without the presence of a solid-liquid interface, the melting of the crystal can proceed only with homogeneous nucleation and hence it becomes possible to super-heat the crystal[30,31]. More generally, melting has been also considered as a process of thermally activated nucleation of the liquid phase inside a superheated crystal[31], with the super-heating effect rising the melting point of a solid typically 20%-30% higher than the experimentally observed one[30]. Another factor, which can favour the super-heating effect in the simulated systems, is the rate, at which the temperatures is changed in the simulations, which usually is many times larger than the rate used by the experimental techniques. It is common to observe melting temperatures or higher *order-disorder* transition temperatures in simulated systems (with different various techniques), when compared to the experiments, as many recent theoretical investigations of crystal structures investigations have reported[26,32]. The solid plastic phase I of cyclopentane in the metastable super-heated state for T > 180K shows a different nature in the orientational dynamical disorder than the same phase in the range 140K-170K, as indicated by the RDF of the system (figures 7.19-7.21).

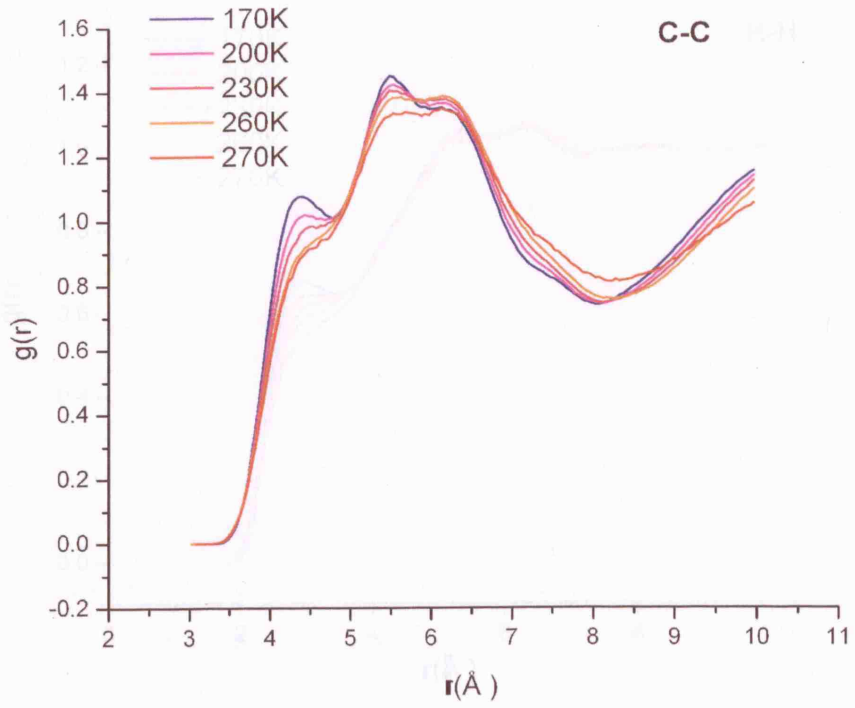


Fig. 7.19: C-C RDF of the MD simulated phase I of cyclopentane at different high temperatures up to the liquid phase at 270K (red curve).

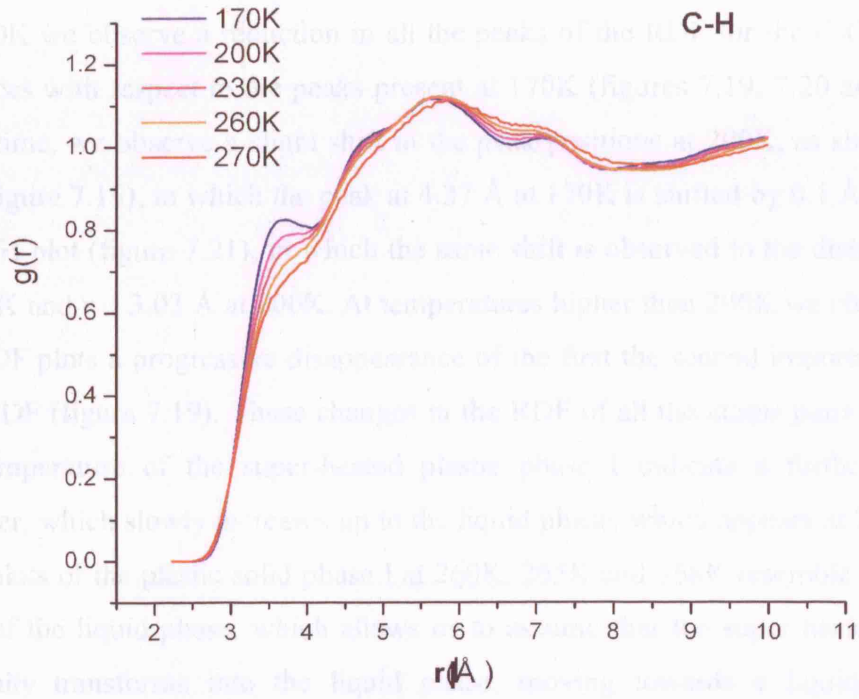


Fig. 7.20: C-H RDF of the MD simulated phase I of cyclopentane at different high temperatures up to the liquid phase at 270K (red curve).

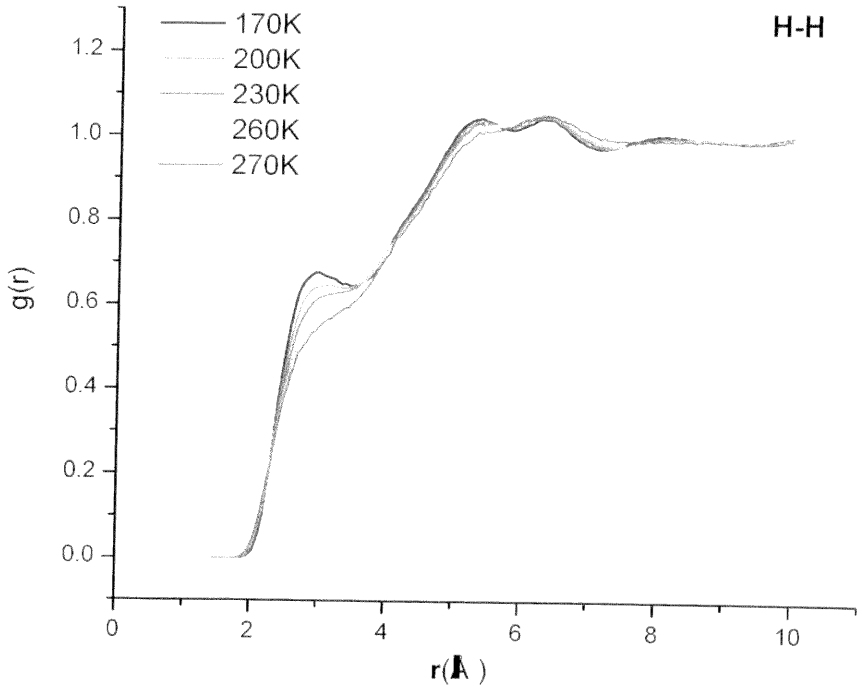


Fig. 7.21: H-H RDF of the MD simulated phase I of cyclopentane at different high temperatures up to the liquid phase at 270K (red curve).

At 200K we observe a reduction in all the peaks of the RDF for the C-C, C-H and H-H distances with respect to the peaks present at 170K (figures 7.19, 7.20 and 7.21) and, at same time, we observe a slight shift in the peak positions at 200K, as shown in the C-C plot (figure 7.19), in which the peak at 4.37 Å at 170K is shifted by 0.1 Å at 200K and in the H-H plot (figure 7.21), in which the same shift is observed to the distance $r \sim 2.93$ Å at 170K and $r \sim 3.03$ Å at 200K. At temperatures higher than 200K we observe in each of the RDF plots a progressive disappearance of the first the second important peaks in the C-C RDF (figure 7.19). These changes in the RDF of all the atoms pairs with increasing the temperature of the super-heated plastic phase I indicate a further increment of disorder, which slowly increases up to the liquid phase, which appears at 270K. All of the RDF plots of the plastic solid phase I at 260K, 265K and 268K resemble very closely the RDF of the liquid phase, which allows us to assume that the super-heated solid phase I gradually transforms into the liquid phase, moving towards a liquid-like molecular structure with increasing temperature.

The presence of the metastable super heated state of the plastic phase I, is confirmed by the mean square displacement (*MSD*) of the system at different temperatures, which

shows a trend typical of a solid even at temperatures higher than 170K (up to 168K). The *MSD* calculation was performed in the NVE ensemble, following the definition, given by the formula [3.56] in section *3.5.4*[13]. At each temperature, the starting configuration of the system, for a simulation in NVE ensemble, was sampled from the trajectory file of the NST simulations and corresponded to the configuration with the cell parameters as close as possible to the average values. We then let the system re-equilibriate in the NVE ensemble for 150ps and finally a production of 400ps was carried out, generating a very dense trajectory file, which reports the configurations of the system every 0.1ps, for a total number of 2500 configurations.

Figure 7.22 shows two different kinds of dependence on time. The first curve corresponds to the solid state phase III at temperature T < 120K, with the time evolution of the *MSD* represented by a straight horizontal line (curve in black in figure 7.22), with a very low value of intercept close to 0 (between ~ 0.19 and ~ 0.32 $Å^2$), which is typical of the *MSD* for solids, often characterised by the presence of oscillations, related to the atomic vibrations of the solid. The second curve at 124K curve in blue is relative to the solid phase III at 124K (blue line in figure 7.22), and shows a linear trend, which is associated to the presence of small molecular diffusion in the system with a diffusion coefficient of ~ $3 \cdot 10^{-14}$ m^2/ps. The very small linear dependence on time of the *MSD* at 124K is also noted in figure 7.23, in which the *MSD* of different solid and plastic phases up to 268K are shown. This slightly linear behaviour of the *MSD* in the solid ordered phase III at such temperature, which is very close to the *order-disorder* transition point, can be explained by considering the necessary re-orientation of the molecules in the system to allow the transition to disorder. This re-orientational motion is never characterised only by pure rotations, and is coupled with a small translational motion to promote the jump to an orientational disorder state. Increment in the translational disorder has been reported in other plastic crystals[3,33].

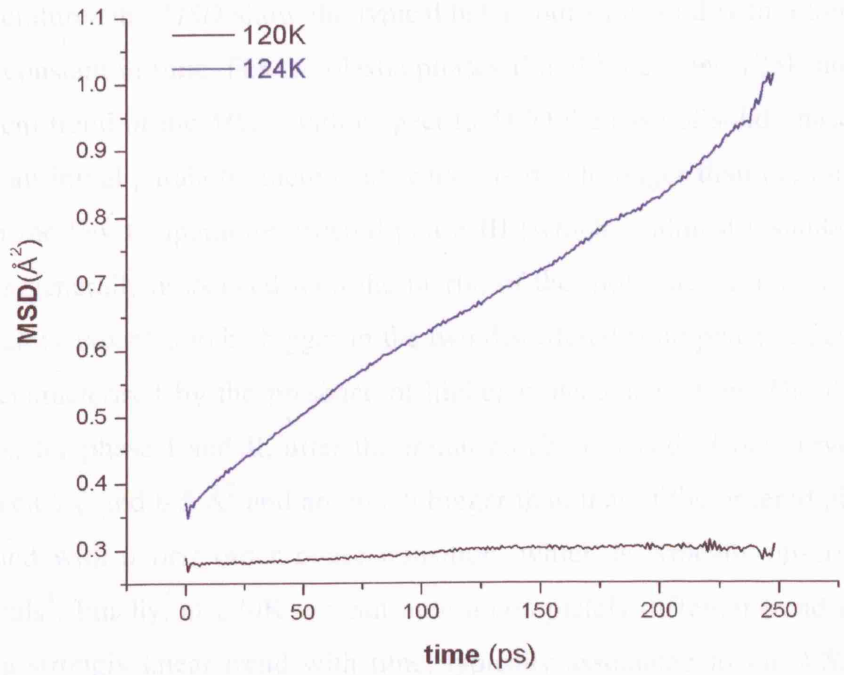


Fig. 7.22: Mean square displacement (*MSD*) of the solid monoclinic phase III of cyclopentane at T < 120K (black) and at 124K (blue, temperature very close to the transition point to the plastic phase II.

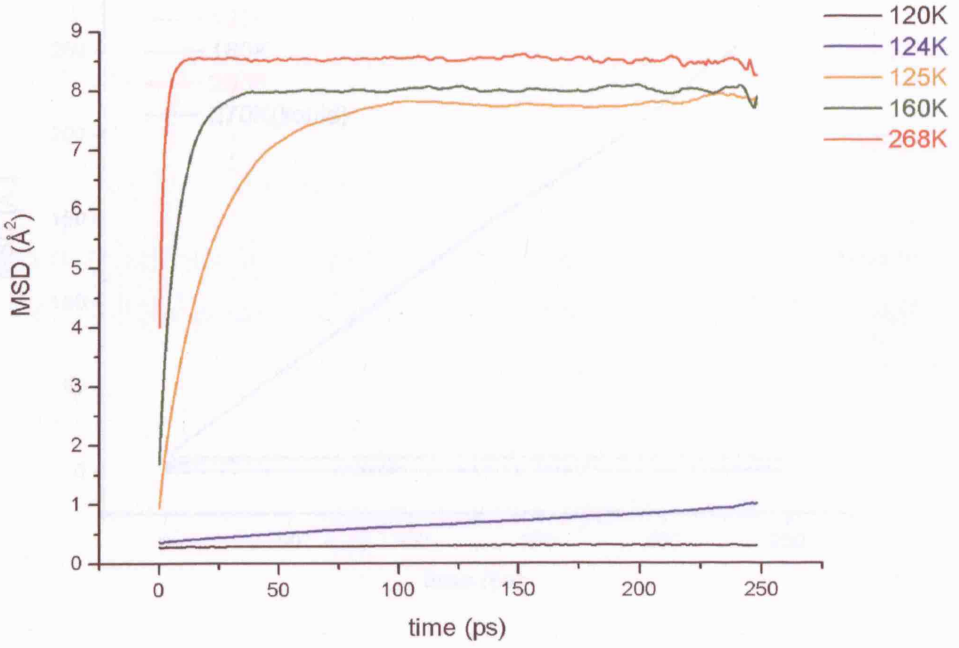


Fig. 7.23: Mean square displacement (*MSD*) of cyclopentane solid and plastic phases at different temperatures: monoclinic solid phase III (black and blue), intermediate plastic phase II (orange), hexagonal plastic phase I (green) and super-heated plastic phase I (red).

At all temperatures the *MSD* show the typical behaviour of a solid with a trend, which is an average constant in time. For the plastic phases II and I, between 125K and 268K, we see a different trend of the *MSD*, with respect to *MSD* the case of solid phase III, (figure 7.21), with an initial parabolic increment, which is much longer than the incipient trend observed in the low temperature ordered phase III (which is almost instantaneous). The increment is generally associated with the inertia of the molecules in the system and the inertia, which is expected to be bigger in the two disordered solid phases of cyclopentane, which are characterised by the presence of higher molecular motion. The final constant *MSD* values, for phase I and II, after the initial parabolic trend of the curve, are in the range between 7.8 and 8.5 $Å^2$ and are much bigger than that of the ordered phase III, and are associated with a first order phase transition, which is typically observed in other plastic crystals[3]. Finally, at 270K we can note a completely different trend in the *MSD*, which has a strongly linear trend with time, typically associated to the *MSD* of liquids (figure 7.24).

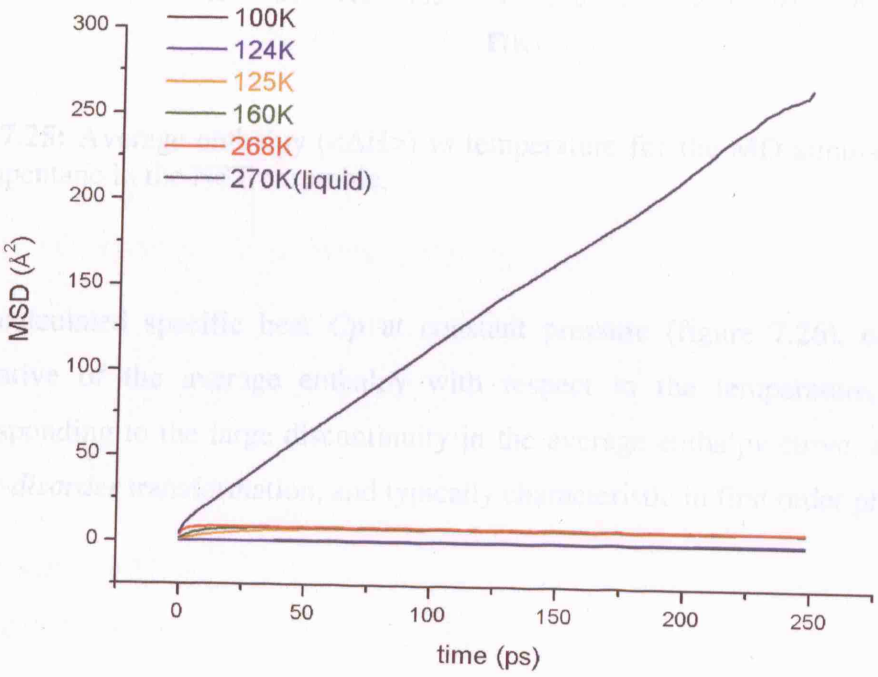


Fig. 7.24: Mean square displacement (*MSD*) of the liquid (dark blue line) and all the solid and plastic phase of cyclopentane.

The nature of the phase transition is more evident by examining the variation in the enthalpy of the system with temperature, reported in figure 7.25, in which we observe a

discontinuity in the average enthalpy of the system at 124K-125K, associated with the *order – disorder* transition from phase III to phase II.

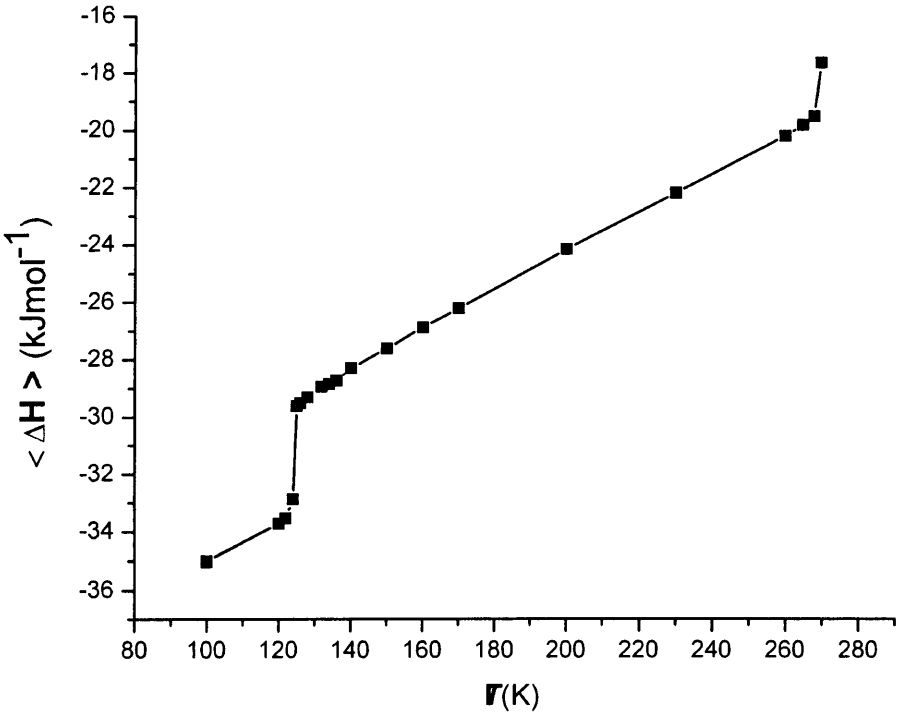


Fig. 7.25: Average enthalpy (<ΔH>) *vs* temperature for the MD simulated structures of cyclopentane in the NST ensemble.

The calculated specific heat *Cp* at constant pressure (figure 7.26), calculated as the derivative of the average enthalpy with respect to the temperature, shows a peak corresponding to the large discontinuity in the average enthalpy curve, associated to the *order-disorder* transformation, and typically characteristic in first order phase transitions.

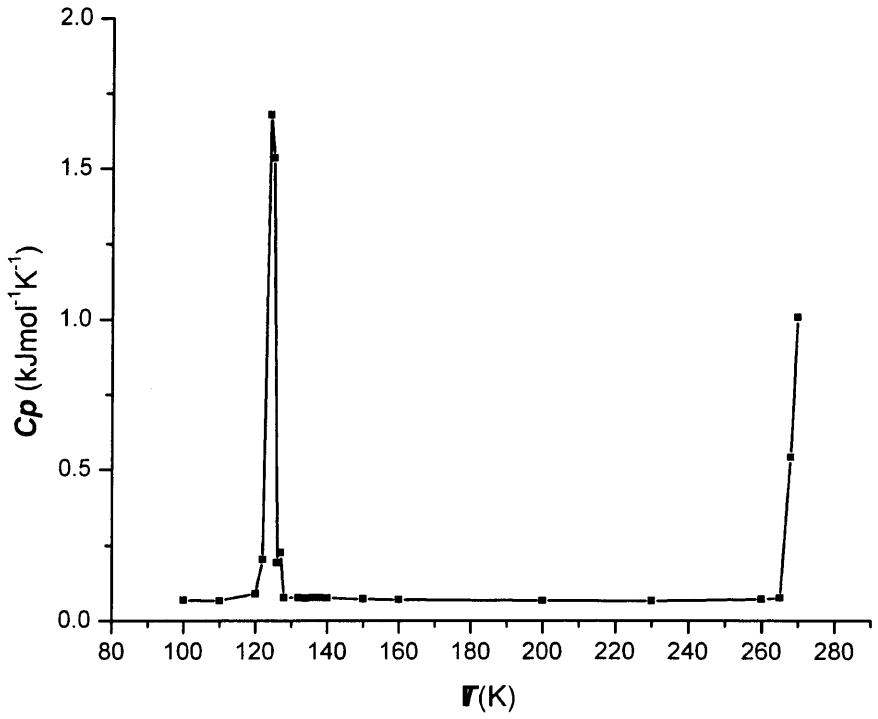


Fig. 7.26: Specific heat at constant pressure (C_p) vs temperature for the MD simulated structures of cyclopentane in anisotropic constant pressure ensemble (NST simulation).

The enthalpy and the entropy of the transition (ΔH_t and ΔS_t) are related to the specific heat *Cp* of the system by the thermodynamics relationships[3]:

$$\Delta S_t = \int_{T_1}^{T_2} \frac{Cp}{T} dT \quad . \qquad [7.4]$$

Using the relationships [7.2] and [7.4], we obtained an approximated estimate of the enthalpy and the entropy of the *order-disorder* transition, by calculating the area under the peaks at 124K-125K respectively in the *Cp* curve and that in the curve *Cp/T vs T*, with values of ΔH_t and ΔS_t for the *order-disorder* transition in cyclopentane respectively of 6.10 kJ mol^{-1} and 0.046 kJ $mol^{-1}K^{-1}$. These calculated values are consistent with experiment[25,34], as summarised in table 7.10.

	$\Delta H_{transition}$ /kJ mol^{-1}	$\Delta S_{transition}$ /J mol^{-1}K^{-1}
Experimental *Ph III → Ph II* *Ph II → Ph I*	4.88 * 0.35*	0.040# 0.0025#
Calculated *Ph III → Ph I*	6.10	0.046

Table 7.10: Experimental and calculated values of the enthalpy and entropy of transition phase III – phase I of cyclopentane. *) The values of experimental enthalpy are taken from the Aston measurements of heat capacity[25]. #) The values of experimental entropy of transformation are taken from Staveley[34].

The large entropy values are usually associated with a large hysteresis in the transition temperature, which were observed in other organic systems, such as quadricyclane[35] and norbornadiene[36].

7.3.5 Simulated annealing of the cyclopentane system from 220K to 20K

Finally we performed an annealing of the system to estimate whether hysteresis occurs on cooling through the order-disorder phase transition. The procedure consists of starting from the high-temperature configuration of the system (268K) and cooling the latter with a small variation in temperature ΔT. For each temperature the system is re-equilibrated at this temperature for a certain period of time. We expect to observe a disorder-order phase transition, starting from the orientational disordered hexagonal super-heated phase I at 220K and cooling down to 20K.This study can give us further information about the real order-disorder transition point and also about the nature of the phase transition. We decrease the temperature of the system with a step ΔT =10K from 220K to 150K, performing NST simulations with a timestep of 0.001ps, for a total simulation time of 1ns, and an equilibration time of 100ps. From 140K to 130K we decrease the temperature of the system more slowly, with a step ΔT = 2K and running longer simulations, up to 2ns, using the same equilibration time and same timestep of the previous simulations between 220K and 150K. Because in the range between 130K and 120K we observed the phase transition during the heating process, in this interval of temperatures we decrease the temperature more slowly, using a smaller step ΔT = 2K; we extended the time of the simulation up to 3ns, with the same equilibration time of 100ps and timestep of 0.001ps.

Finally from 110K to 20K we used again a step ΔT =10K on decreasing the temperature of the system, running simulations for 1ns, with the same timestep and the same equilibration time, which was used in the simulations at the previous temperatures. In all the simulations performed in the cooling we used a cutoff of 10Å for the repulsive-dispersion contribution, simulated with the same interatomic potential used in the heating process. A precision of 10^{-6} is used in the calculation of the electrostatics contribution by the Ewald summation and a cutoff of 15 Å is used in the latter for the sum in the real space. The results for the variation of the lattice parameters and the volume of the unit cell of the structure, as function of temperature are reported in table 7.11 and in the corresponding plots are shown in figure 7.27 and figure 7.28, together with the data obtained from the simulated heating process.

T (K)	*a* (Å)	*b* (Å)	*c* (Å)	*α* (°)	*β* (°)	*γ* (°)	*Vol/Z* (Å^3)
220	9.716	6.035	10.460	89.95	90.06	90.00	153.27
210	9.687	6.014	10.442	89.98	90.09	90.10	152.04
200	9.652	6.009	10.407	90.04	90.03	89.99	150.84
190	9.634	5.984	10.380	89.93	90.00	90.14	149.56
180	9.602	5.976	10.346	90.00	90.03	90.00	148.37
170	9.581	5.950	10.329	89.99	89.97	90.09	147.16
160	9.551	5.934	10.303	90.06	89.96	90.09	145.93
150	9.528	5.913	10.276	90.16	89.93	90.07	144.69
148	9.532	5.912	10.253	89.92	90.00	90.07	144.41
146	9.525	5.905	10.257	89.93	90.05	89.95	144.19
144	9.523	5.905	10.244	89.96	90.04	90.06	143.96
142	9.520	5.904	10.229	90.08	90.04	90.07	143.69
140	9.518	5.905	10.213	90.11	89.98	89.98	143.44
138	9.527	5.889	10.211	89.96	90.03	90.01	143.17
136	9.507	5.894	10.209	89.81	89.98	90.07	142.97
134	9.520	5.866	10.226	89.84	90.04	89.81	142.72
132	9.505	5.874	10.206	89.94	89.89	90.05	142.42
130	9.493	5.880	10.194	90.06	89.91	90.07	142.20
128	9.518	5.856	10.188	89.86	89.93	90.01	141.93
127	9.497	5.869	10.181	90.05	90.04	90.10	141.80
126	9.502	5.855	10.189	90.01	90.05	89.98	141.68
125	9.489	5.865	10.179	90.11	89.99	90.05	141.59
124	9.495	5.855	10.179	89.98	89.96	90.06	141.44
122	9.491	5.851	10.171	90.16	89.96	90.00	141.17
120	9.498	5.848	10.152	89.97	90.10	89.82	140.92
110	9.481	5.817	10.121	89.48	90.11	90.34	139.49
100	9.588	5.782	9.9587	90.97	89.81	90.36	137.97
90	9.493	5.708	10.080	90.58	90.34	90.68	136.50
80	9.524	5.653	10.070	90.42	89.82	90.19	135.53
70	9.558	5.674	9.928	89.25	89.81	90.52	134.57
60	9.538	5.682	9.879	89.31	90.38	90.71	133.82
50	9.514	5.674	9.866	89.32	90.30	90.71	133.14
40	9.487	5.664	9.866	89.41	90.32	90.49	132.54
30	9.474	5.657	9.853	89.41	90.31	90.47	132.00
20	9.461	5.650	9.840	89.42	90.30	90.46	131.49

Table 7.11: Structural parameters as function of temperatures for cyclopentane crystal structure, in the MD simulated cooling.

From the structural data reported in table 7.11 is evident the irreversibility of the simulated heating process for the β angle, which does not revert to the values of 76-75°, characterising the monoclinic symmetry of the unit cell at temperatures lower than 124K. Indeed, at these temperatures, in the cooling process, the β angle remains fixed at ~ 90.0°,

with the system preserving orthorhombic symmetry of phase I. The other lattice parameters decrease with temperature up to at most 0.2 Å for the ***a*** axis, 0.4 Å for the ***b*** axis and 0.6 Å for the ***c*** axis, while the other two cell angles α and γ fluctuate around 90.0°. The contrasts between the heating and cooling curves are shown in figures 7.27-7.32.

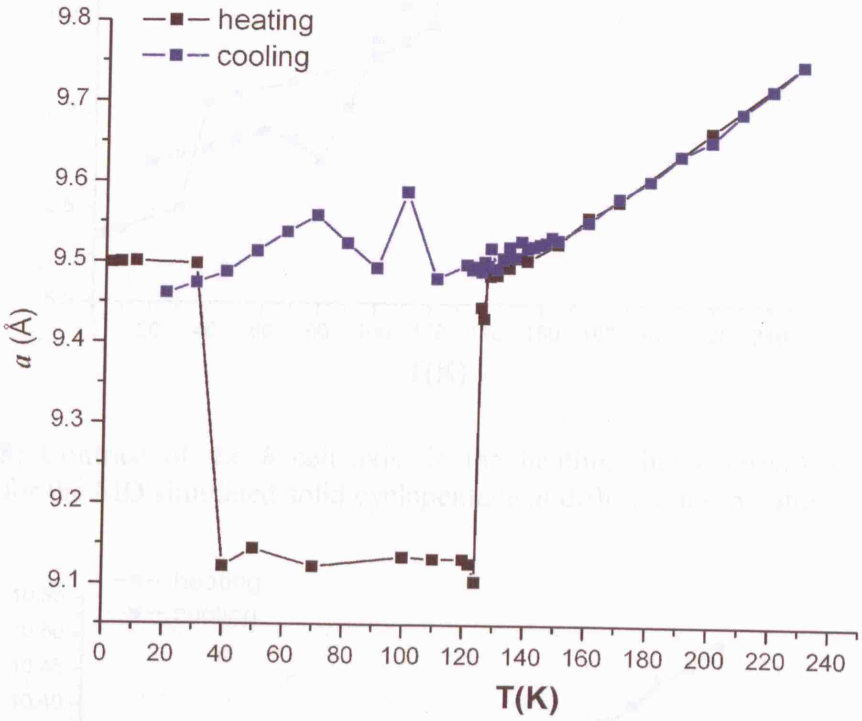


Fig.7.27: Contrast of the ***a*** cell axis, in the heating (black curve) and cooling (blue curve), for the MD simulated solid cyclopentane at different temperature.

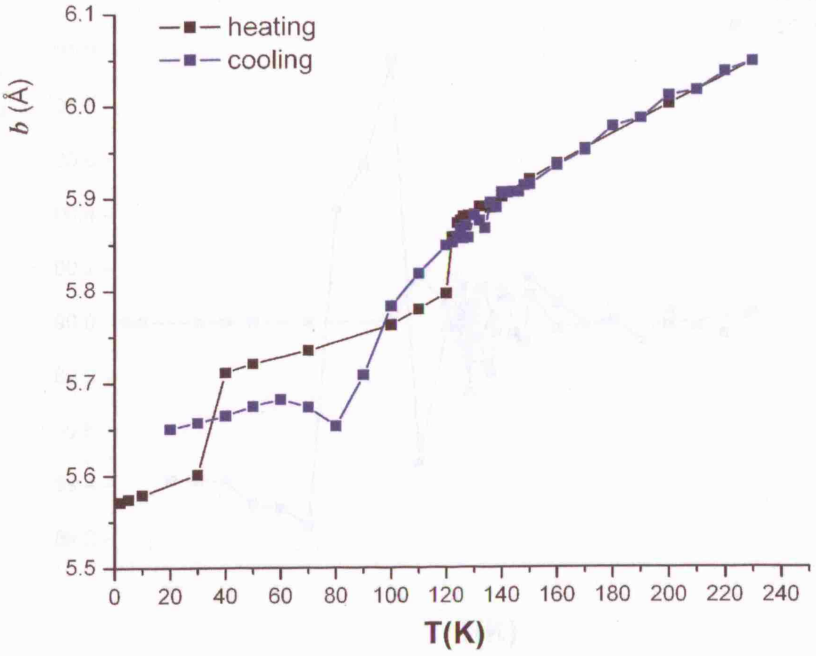


Fig.7.28: Contrast of the ***b*** cell axis, in the heating (black curve) and cooling (blue curve), for the MD simulated solid cyclopentane at different temperatures.

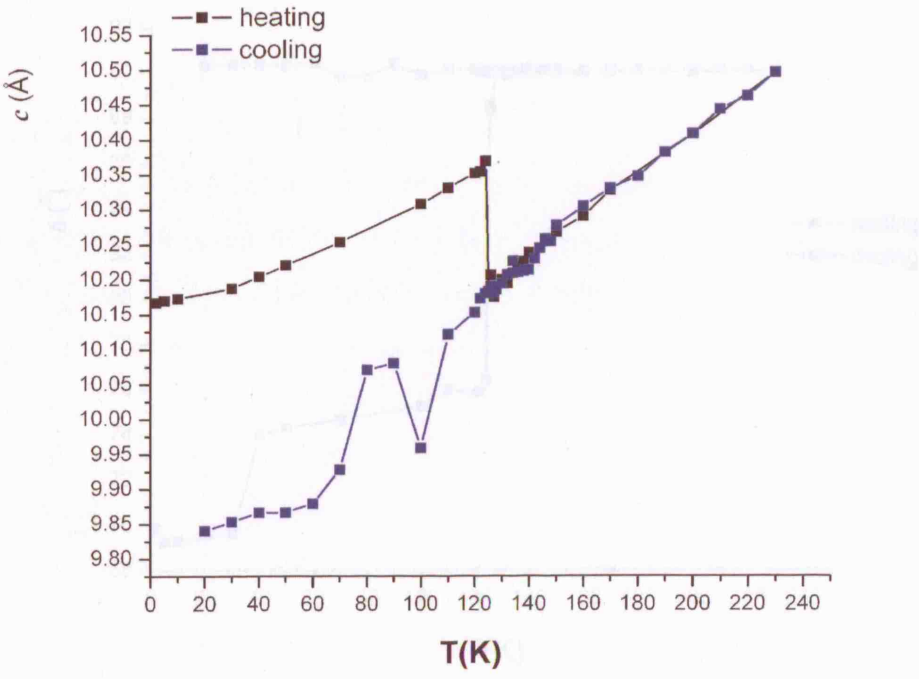


Fig.7.29: Contrast of the ***c*** cell axis, in the heating (black curve) and cooling (blue curve), for the MD simulated solid cyclopentane at different temperatures.

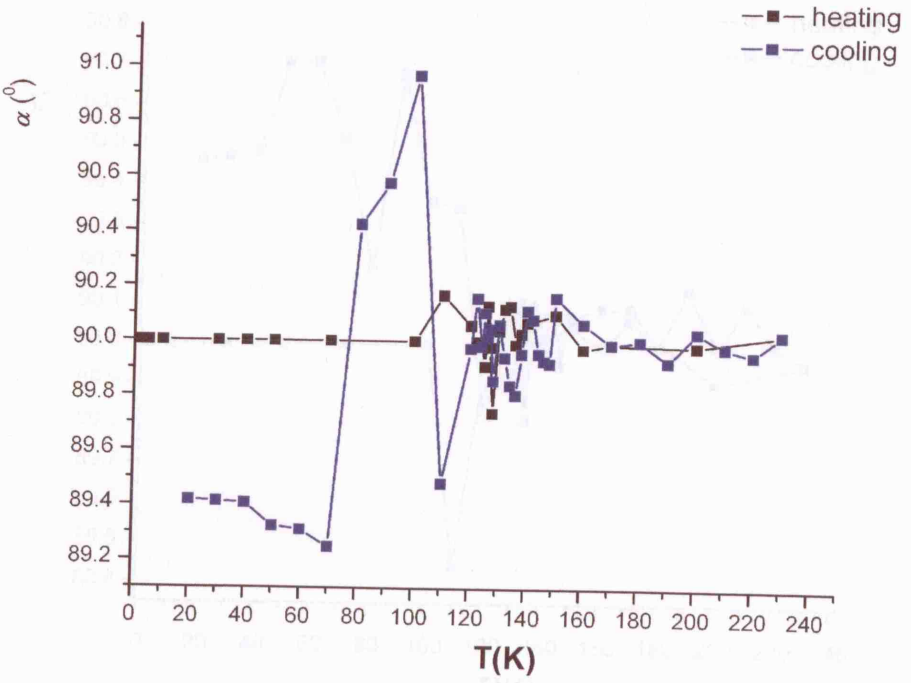


Fig.7.30: Contrast of the α cell angle, in the heating (black curve) and cooling (blue curve), for the MD simulated solid cyclopentane at different temperatures.

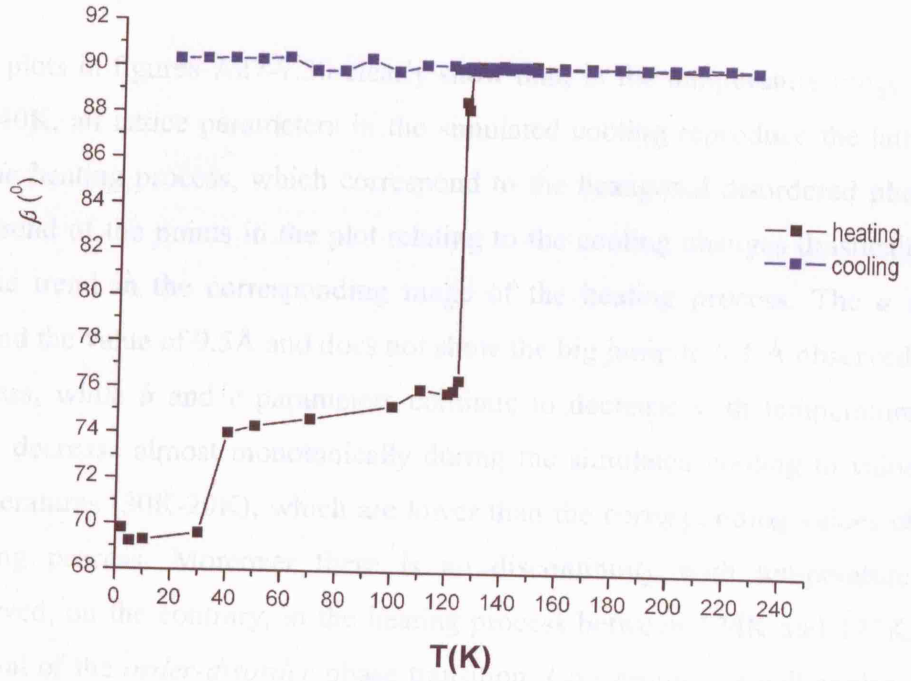


Fig.7.31: Contrast of the β cell angle, in the heating (black curve) and cooling (blue curve), for the MD simulated solid cyclopentane at different temperatures.

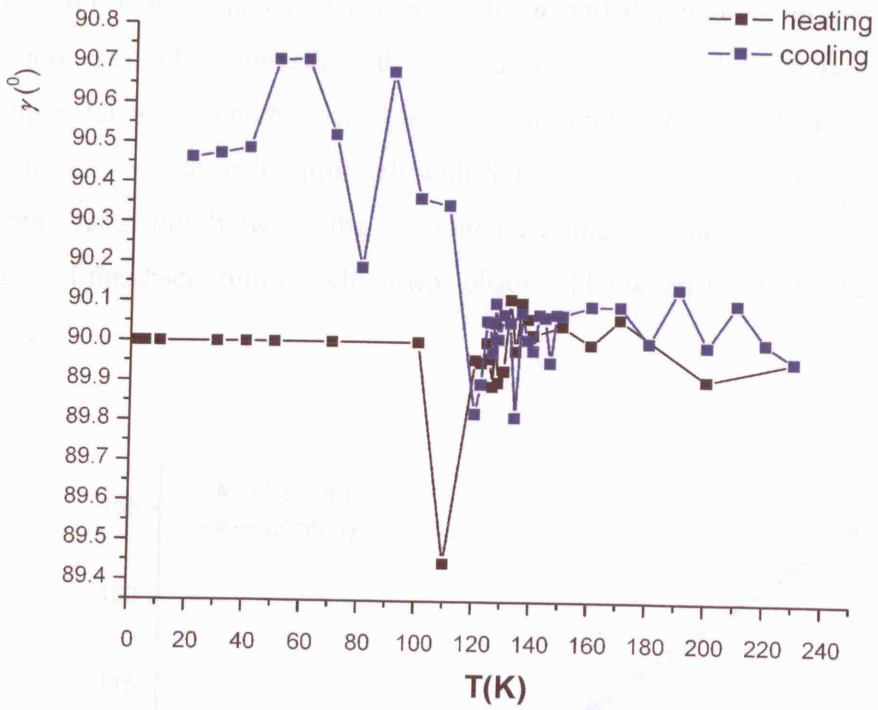


Fig.7.32: Contrast of the γ cell angle, in the heating (black curve) and cooling (blue curve), for the MD simulated solid cyclopentane at different temperatures.

The plots in figures 7.27-7.30 clearly show that, in the temperature range from 220K up to 140K, all lattice parameters in the simulated cooling reproduce the lattice parameters in the heating process, which correspond to the hexagonal disordered phase I. At 130K the trend of the points in the plot relating to the cooling changes drastically with respect to the trend in the corresponding range of the heating process. The ***a*** axis fluctuates around the value of 9.5Å and does not show the big jump to 9.1 Å observed in the heating process, while ***b*** and ***c*** parameters continue to decrease with temperature. Both of the latter decrease almost monotonically during the simulated cooling to values at very low temperatures (30K-20K), which are lower than the corresponding values observed in the heating process. Moreover there is no discontinuity with temperature, which was observed, on the contrary, in the heating process between 124K and 127K, which is the interval of the *order-disorder* phase transition. Concerning the cell angles, as previously discussed, we note (figure 7.31) a "freezing" of β angle at ~ 90°, which can not revert to the values of 74-75° observed in the monoclinic ordered phase III under simulated heating. The behaviour of the other two angles, α and γ is also interesting. In the phase I range of temperatures they fluctuate around 90° like those in the simulated heating, but at

low temperatures T ≤ 120K, they are subjected to even bigger fluctuations which promote changes up to 91.0° and down to 89.2° for α and fluctuations around 90.5° for γ. This behaviour strongly contrasts with that of the corresponding angles in the simulated heating process, in which both of are constant and stable at 90°. Finally, looking at the trend in the volume of the unit cell with temperature (figure 7.33), we observe a clearly different behaviour between the simulated cooling and the heating processes, with the absence of the discontinuity, which was observed between 124K and 127K, in the heating process.

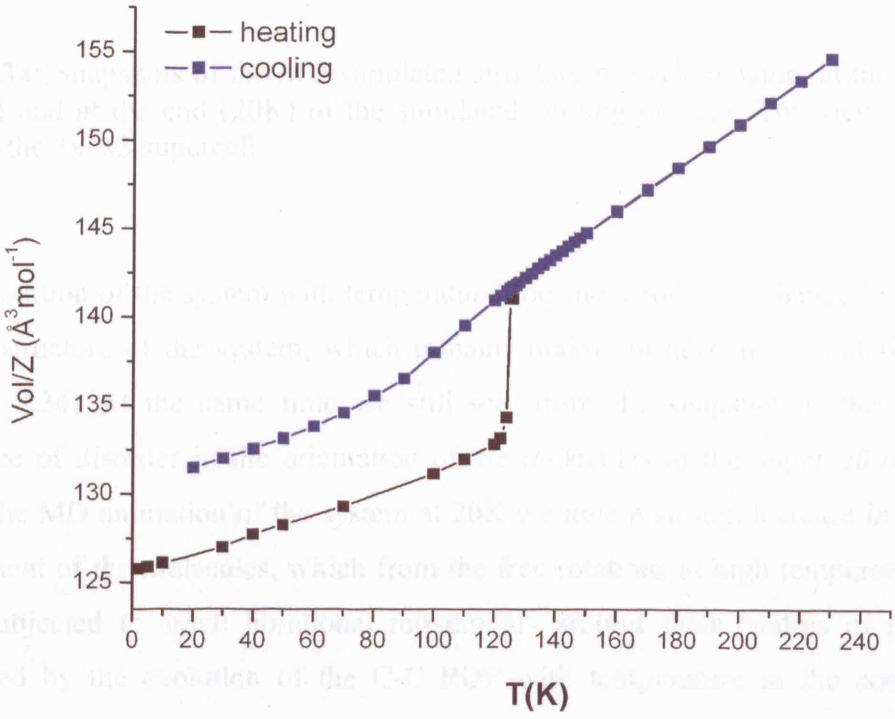


Fig. 7.33: Trend in the volume of the solid cyclopentane unit cell with temperature, in the simulated cooling (blue curve) and in the simulated heating (black curve).

Under simulated cooling (blue curve in figure 7.33) the volume decreases linearly with temperature from 220K to 128K, following the trend of the points obtained in the heating process, but for T < 128K continues to decrease almost linearly, with a smooth inflection point in the range between 110K and 80K, moving away from the discontinuous behaviour occurring in the reverse heating process with a final value at low-temperature, which is ~ 6Å^3/mol larger than that observed at the same temperatures during the heating. Figure 7.34 shows some snapshots of the cyclopentane system, at the beginning and at the end of its evolution during the simulated cooling.

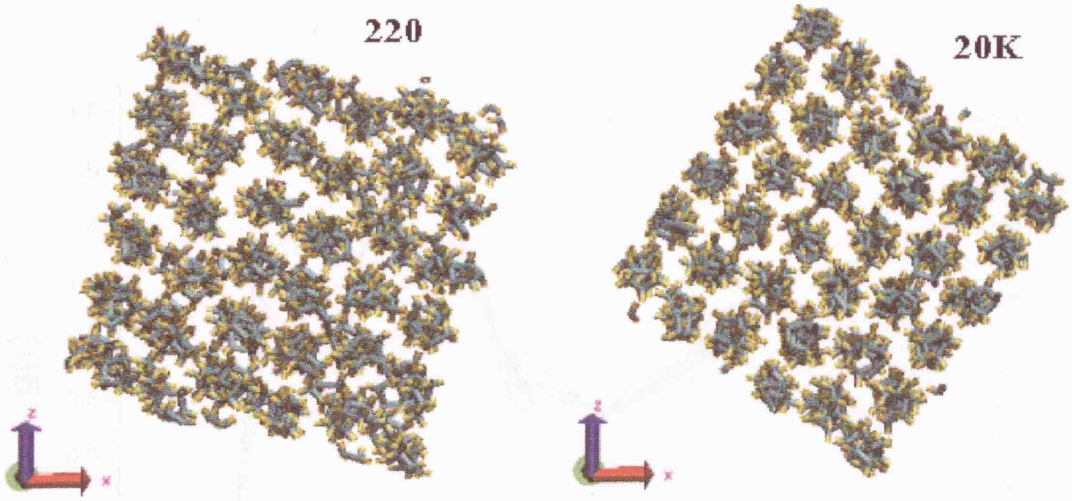


Fig. 7.34: Snapshots of the MD simulated structure of cyclopentane, at the starting point (220K) and at the end (20K) of the simulated cooling process. The view is along the ***b*** axis of the 3x5x3 supercell.

The evolution of the system with temperature does not produce a change in the symmetry in the structure of the system, which remains mainly orthorhombic and slightly rotated (figure 7.34) At the same time we still see, from the snapshot of the structure, the presence of disorder in the orientation of the molecules in the supercell of the system. From the MD animation of the system at 20K we note a strong decrease in the rotational movement of the molecules, which from the free rotations at high temperatures, are now only subjected to small librational movements around their centres of mass, as also indicated by the evolution of the C-C RDF with temperature in the cooling process, reported in figure 7.35.

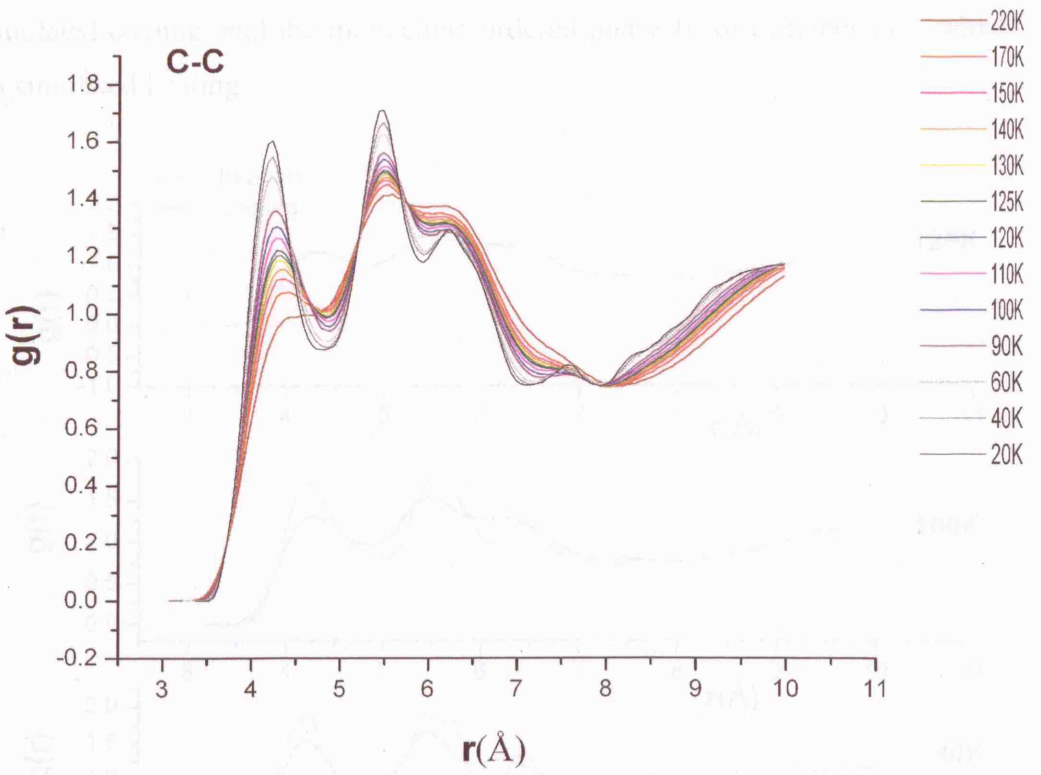


Fig. 7.35: Evolution of the C-C RDF with temperature in the MD simulated cooling of the cyclopentane system in the NST ensemble, from 220K to 20K.

In the C-C RDF profile, we observe that the main peak at 4.21Å, when the system is at 20K, increases by a small amount with raising the temperatures up to ~ 160K and results shifted of few decimal places at temperatures up to160K. For T ≥ 170K the peak disappears. The other peaks at 20K (5.46Å, 6.21Å and 7.54Å) already disappear at 90K. This variation in the RDF profiles, on cooling the system, reflects an increase of the order in the system, which does not concern an increase in the number of iso-oriented the molecules (as in figure 7.6a), but solely a decrease in the rotational motion of the molecules, which do not rotate around their centres of mass as previously discussed. Indeed, the few molecular orientations characteristic of the ordered monoclinic phase III in a low temperature stable state (figure 7.6 a), are now changed in a multitude of different possible orientations with respect to the cell axes. The system can be viewed as in a metastable "glassy" state, in which the orientational disorder of the molecules in the high temperature phase I has been frozen. Hence this state can be seen as a sort of super-cooled phase I. The contrast between the C-C RDFs at three different temperatures (125K, 100K, 20K), for the heating and cooling processes (figure 7.36), confirms the

structural difference between the metastable phase, obtained at low temperatures from the simulated cooling, and the monoclinic ordered phase III of cyclopentane, which is stable on simulated heating.

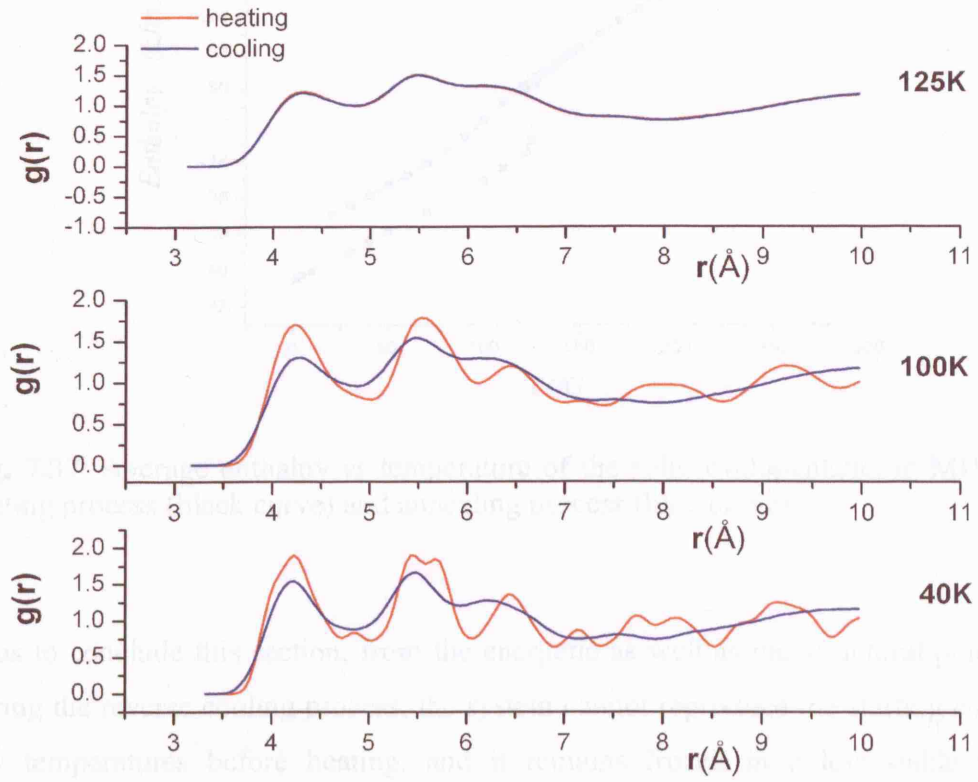


Fig.7.36: Contrast of the C-C RDF profiles of the MD simulated cyclopentane in the heating (red curve) and in the cooling (blue curve) process, at three different temperatures.

From the comparison of the trend of the average enthalpies with temperature, for the MD simulated system in the heating (figure 7.37: black curve) and cooling (figure 7.37: blue curve), at T =127K we do not observe a steep slope in the trend of the average enthalpies in the cooling process, which occurs, on the contrary, in the heating process. From high temperatures, up to 128K, the two processes show reversibility, with the two plots running parallel to each other, but at T < 128K, the enthalpies in the simulated cooling remains at higher values compared to those on the heating.

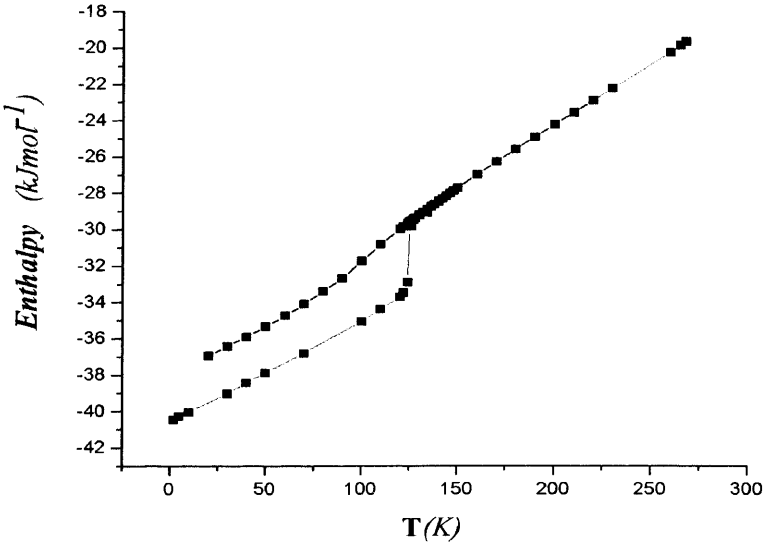


Fig. 7.37: Average enthalpy *vs* temperature of the solid cyclopentane, in MD simulated heating process (black curve) and annealing process (blue curve).

Thus to conclude this section, from the energetic as well as the structural point of view, during the reverse cooling process, the system cannot reproduce the starting conditions at low temperatures before heating, and it remains frozen in a less stable state. This characteristic of the system, similar to a hysteresis, which is typical of first order phase transitions, has been observed in many different systems, on varying temperature[36,35] or pressure[37]. It is well established[35] that the hysteresis is generated by a big change in the transformation entropy in the order-disorder phase transition, which is likely to be associated with the requirement for nucleation and growth of the new within the old structure. This process, in fact, generates large strains in the crystal structure, which are responsible of the hysteresis in the transition temperature. However in this system we might not observe a complete hysteresis, but simply a failure of the system in reproducing the initial conditions at low temperatures after simulated heating and cooling. This behaviour could be due to two different factors. The first concerns the limited dimension of the system (a 3x5x3 supercell, with 180 molecules), on which the periodic boundary conditions are applied, which can affect the ability of the system to regain the initial order in the configurations and orientations in the low temperature structure. The second concerns the cooling protocol, in which the use of an initial step of 10K, in decreasing the temperature, might cooled the system so rapidly that prevented a proper re-arrangement of the molecules in the system at lower temperature. The

molecular re-orientation into an iso-oriented, ordered system needs probably much smaller temperature decrease (ΔT), longer equilibration periods and longer simulations in time in the cooling process, which would imply very large computation cost. Nevertheless it is quite likely that on increasing the system size and decreasing ΔT there will still be substantial hysteresis in system, which will regenerate a considerably different ordered monoclinic phase III. The restoring of the order into the molecular orientation of a disordered system possibly requires a nucleation process, with the presence of small ordered *nuclei*, which can lead and help the cooperative re-orientation of the molecules. MD simulations are far from being able to model such kind of processes, without the explicit inclusion of a surface, or nuclei inside the system. This situation can be schematised from an energetic point of view by a big energetic barrier, which needs to be overcome by the system in order to transit from the disordered phase I to the low temperature order phase III (see the scheme in figure 7.38). MD methods are not suitable to model events which require the overcoming big energetic barriers.

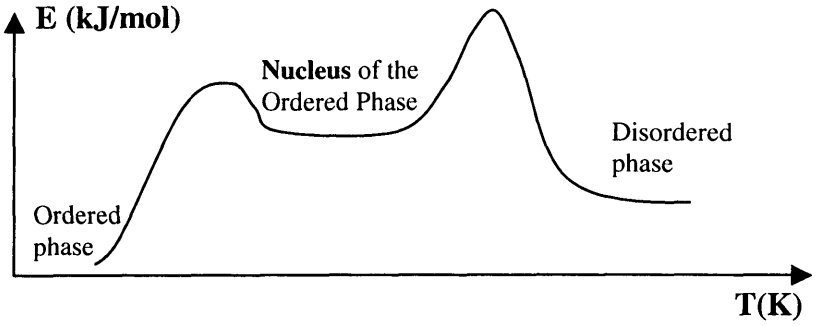


Fig. 7.38: Schematic illustration of a possible energetic landscape, involved in the *disorder-order* phase transition, in the cooling process of the solid cyclopentane.

This discussion is an aspect of a much larger, important and still debated issue: do second order phase transitions exist, or are all phase transitions first order, in the sense that all of them occur through mechanism of nucleation and growth?[38]. A final important consideration is that the MD simulation effectively simulate a "glassy phase" of cyclopentane, which has not been observed yet in experiments. Glassy crystals have been observed at low temperature, from a rapid cooling process of a plastic phase and have been reported in the literature, which describe them as fragile metastable structures[39], characterised by a particular local order symmetry[40].

7.3.6 X-ray Powder Pattern Calculation

The experimental investigation of the cyclopentane crystal phases performed by Dr. C.K. Leech at Rutherford Appleton Laboratory through X-Ray analysis with a monocromatic radiation having λ = 0.8002 Å, showed the presence of three distinct powder patterns, corresponding to three different phases[2]. The monoclinic ordered phase III at low temperature was fully indexed and corresponds to the structure solved by R. Bose[2]. The diffraction patterns relating to this phase are observed in a range of 20K, from 100K to 120K and are characterised by a diffraction pattern with many peaks (figure 7.39a, blue trace). At 120K, a different powder pattern appears (figure 7.40a, green curve), with the presence of many dense peaks around the value 2θ =10.33°. The data are not easy to resolve, because of the presence of a pronounced curvature of the background line, which is a signal of presence of disorder in the structure. An enlargement of the region at low angles (figure 7.40b) shows the presence of a very weak peaks around 2θ = 6°, indicating a big unit cell for the partially disordered phase II. At a temperature of 138K, the X-Ray pattern clearly switches to a third distinct shape (figure 7.41a, red trace), characterised by a large feature in the background, typical of a very disordered system and by the presence of few peaks, characteristic of an hexagonal symmetry of the unit cell, previously described. This kind of pattern remains unaltered for higher temperatures up to 144K.

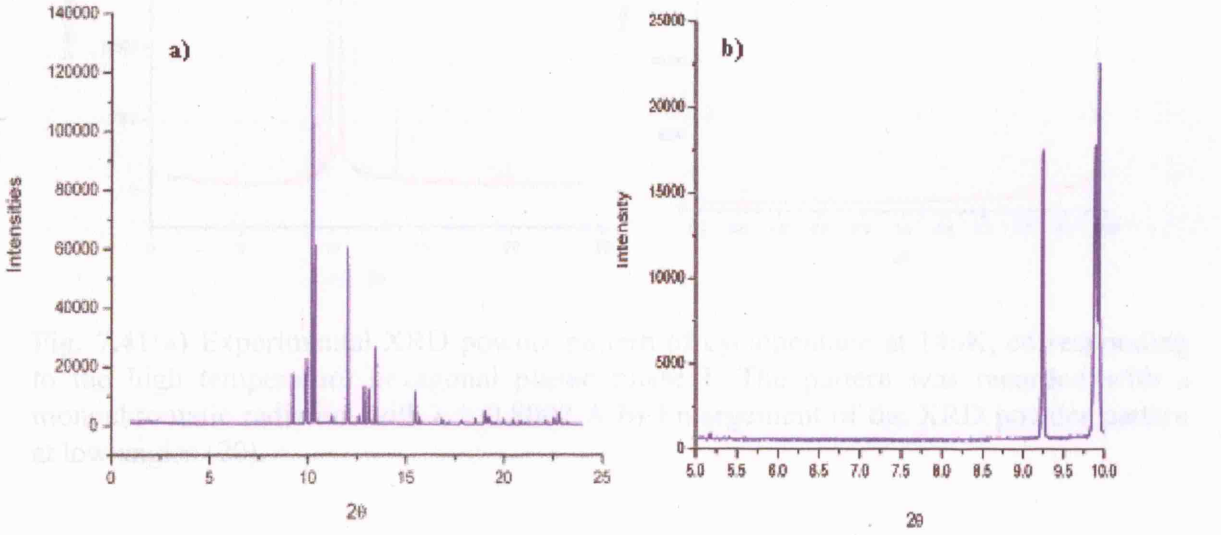


Fig. 7.39:a) Experimental XRD powder pattern of cyclopentane at 100K, corresponding to the monoclinic ordered phase III. The pattern was recorded with a monochromatic radiation with λ = 0.8002 Å **b)** Enlargement of the XRD powder pattern at low angles (2θ).

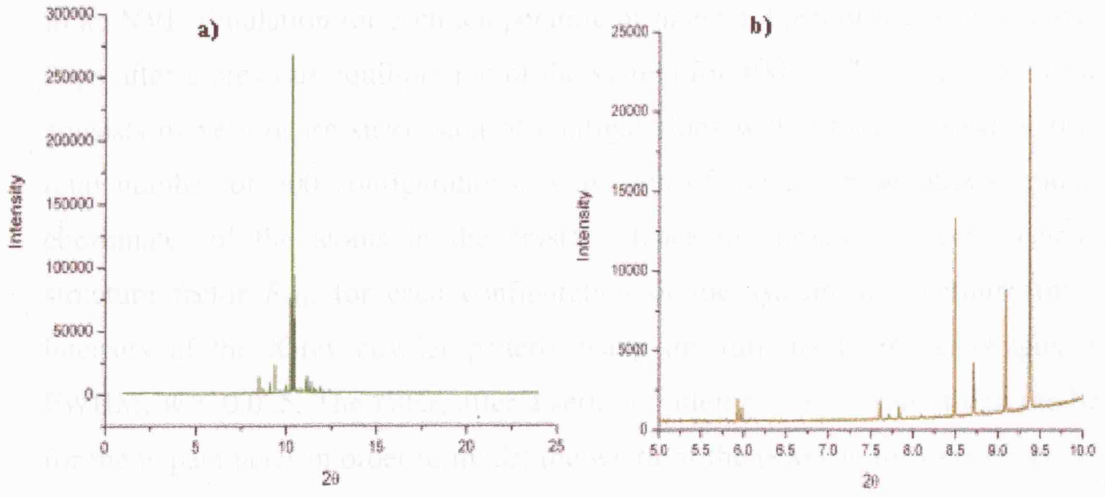


Fig. 7.40: a) Experimental XRD powder pattern of cyclopentane at 126K, corresponding to the intermediate plastic phase II. The pattern was recorded with a monochromatic radiation with $\lambda = 0.8002$ Å **b)** Enlargement of the XRD powder pattern at low angles (2θ).

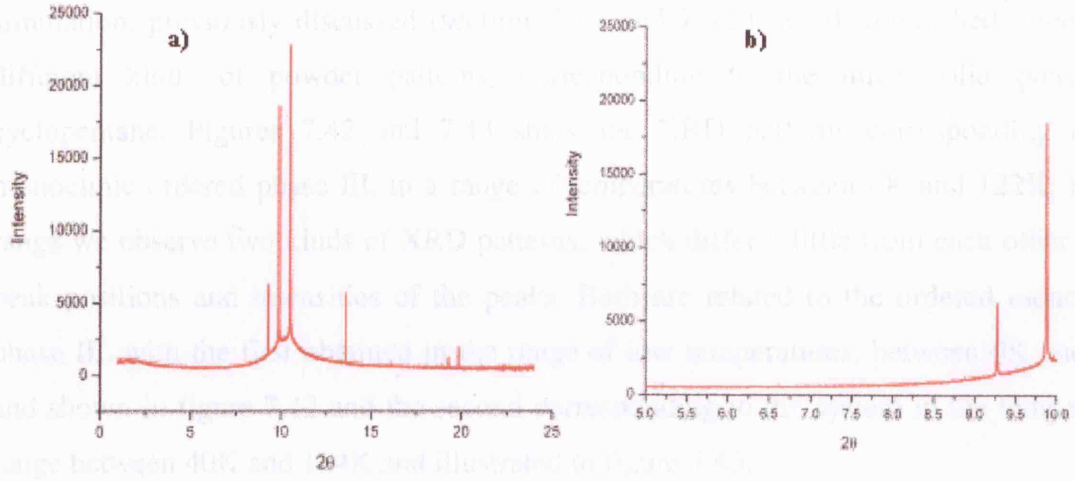


Fig. 7.41:a) Experimental XRD powder pattern of cyclopentane at 140K, corresponding to the high temperature hexagonal plastic phase I. The pattern was recorded with a monochromatic radiation with $\lambda = 0.8002$ Å **b)** Enlargement of the XRD powder pattern at low angles (2θ).

X-Ray powder patterns calculation of the MD simulated structures at different temperatures was performed using a program written by Dr. M. Leslie and Dr. P. Karamertzanis. In order to avoid fluctuations in the lattice parameters, which can strongly affect the calculation of the structure factor, we had to gather data about the atomic positions in the NVE ensemble. Hence, a snapshot of the simulation with cell parameters

as close to the average cell dimensions as possible, was used for the initial configuration in an NVE simulation for each temperature of interest. Each of the NVE simulations last 20ps after a previous equilibration of the system for 100ps. The trajectory of the system consists of very dense succession of configurations with a time interval of 0.1ps, for a total number of 200 configurations, with each of these configurations containing the coordinates of the atoms in the crystal. Hence the program directly calculates the structure factor F_{hkl}, for each configuration of the system at a certain time and the intensity of the X-ray powder pattern, using the formula [3.76] in chapter 3 with a FWHM, w = 0.025. The latter, after a serial of attempts, was found to be the best value for the w parameter in order to model the width of the peaks in the experimental patterns as accurately as possible. The program adopts the same value of wave-length λ = 0.8002 Å of the radiation used in the experimental analysis. Finally it performs an average of the integrated intensities of the calculated patterns over all the configurations of the trajectory of the system for the specific temperature.

Coherently with the trend of the structural parameters with temperature in the MD simulation, previously discussed (section *7.3.2* and *7.3.3*), we distinguished three main different kinds of powder patterns, corresponding to the three solid phases of cyclopentane. Figures 7.42 and 7.43 show the XRD patterns corresponding to the monoclinic ordered phase III, in a range of temperatures between 0K and 122K. In this range we observe two kinds of XRD patterns, which differ a little from each other in the peak positions and intensities of the peaks. Both are related to the ordered monoclinic phase III, with the first obtained in the range of low temperatures, between 0K and 35K and shown in figure 7.42 and the second corresponding to the system in the temperature range between 40K and 124K and illustrated in figure 7.43.

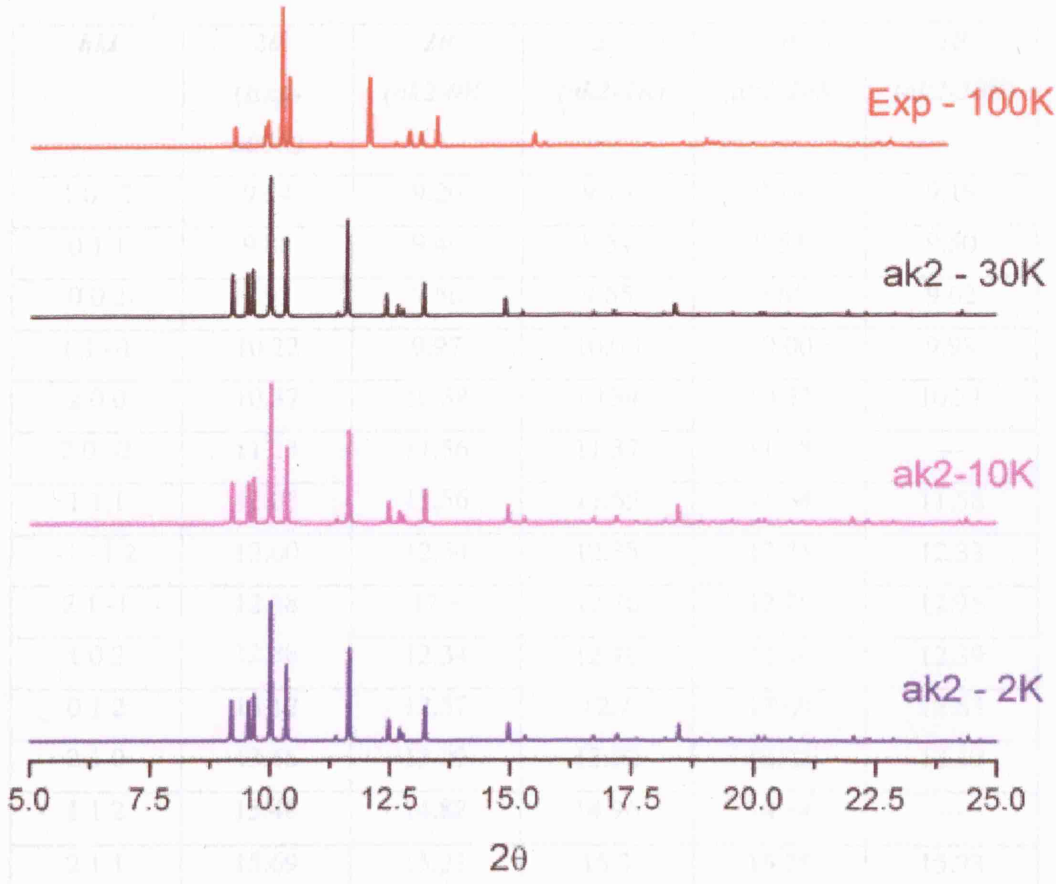


Fig. 7.42: Calculated XRD powder patterns of the MD simulated phase III (*ak2*) in the low temperature range between 0K and 35K, compared to the experimental XRD powder pattern at 100K (*Exp* in red).

In table 7.12 the peak positions corresponding to the main [*hkl*] reflections present in the patterns for the very low temperature monoclinic phase III and in table 7.13 we report integrated intensities of the reflections.

hkl	*2θ (Exp-100K)*	*2θ (ak2-0K)*	*2θ (ak2-2K)*	*2θ (ak2-10K)*	*2θ (ak2-30K)*
1 0 –2	9.24	9.20	9.19	9.19	9.19
0 1 1	9.88	9.46	9.54	9.53	9.50
0 0 2	9.93	9.56	9.65	9.65	9.62
1 1 –1	10.22	9.97	10.00	10.00	9.98
2 0 0	10.37	10.38	10.34	10.33	10.31
2 0 –2	11.23	11.56	11.37	11.38	---
1 1 1	12.03	11.56	11.65	11.64	11.58
-1 –1 2	12.60	12.34	12.35	12.35	12.33
2 1 -1	12.88	12.8	12.76	12.76	12.75
1 0 2	12.88	12.34	12.48	12.46	12.39
0 1 2	13.12	12.57	12.7	12.69	12.63
2 1 0	13.45	13.22	13.22	13.22	13.19
1 1 2	15.48	14.82	14.96	14.94	---
2 1 1	15.69	15.21	15.3	15.28	15.23
-2 –1 3	16.84	16.73	-------	16.71	16.71
-3 –1 2	17.17	17.34	-------	17.19	17.23
3 1 0	17.80	17.62	17.61	17.59	17.55
0 2 1	17.80	17.06	17.21	17.19	17.13
1 0 –4	18.54	18.07	18.19	18.18	18.14
2 0 –4	19.54	18.46	-------	-------	18.44
-3 –1 3	18.94	------	19.00	19.00	19.02
2 1 2	19.02	------	-------	18.45	---
1 2 1	19.14	18.32	18.47	-------	18.37
4 0 –2	19.25	19.77	19.54	19.54	19.56
1 2 –2	-------	18.81	18.93	18.91	18.85
0 2 2	-------	19	19.16	19.14	19.07

Table 7.12: Survey of the main Bragg reflections and peak positions 2θ, for the experimental (*Exp-100K*) and MD simulated structures (*ak2*) of cyclopentane in the low temperature range, between 0K and 35K, corresponding to phase III. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns.

hkl	*IntegInt (Exp-100K)*	*IntegInt (ak2-0K)*	*IntegInt (ak2-2K)*	*IntegInt (ak2-10K)*	*IntegInt (ak2-30K)*
1 0 –2	0.0953	0.318	0.279	0.285	0.289
0 1 1	0.152	0.325	0.281	0.288	0.296
0 0 2	0.154	0.315	0.321	0.326	0.322
1 1 –1	1	1	1	1	1
2 0 0	0.154	0.569	0.544	0.538	0.556
2 0 –2	0.0319	0.727	0.0302	0.0302	---
1 1 1	0.926		0.658	0.662	0.656
-1 –1 2	0.0438	0.191	0.0132	0.0127	0.0150
2 1 -1	0.163	0.0541	0.0508	0.0500	0.0498
1 0 2		0.191	0.146	0.148	0.150
0 1 2	---	0.0737	0.0760	0.0761	0.0749
2 1 0	0.213	0.248	0.238	0.233	0.237
1 1 2	0.409	0.134	0.123	0.123	---
2 1 1	0.329	0.035	0.0294	0.0306	0.0299
-2 –1 3	0.0709	0.0368	---	0.0333	0.0342
-3 –1 2	0.0594	0.00969	---	0.0479	0.00987
3 1 0	0.0315	---	0.0105	0.0102	0.00983
0 2 1	0.0522	0.0404	0.0486	0.0479	0.0381
1 0 –4	---	0.0168	0.0154	0.0160	0.0166
2 0 –4	0.0737	0.0215	---	---	0.0185
-3 –1 3	---	---	0.00266	0.00279	0.00253
2 1 2	0.0144	---	---	0.134	---
1 2 1	0.234	0.124	0.134	---	0.115
4 0 –2	0.142	0.0168	0.0210	0.0212	0.0183
1 2 –2	-----	0.0123	0.0104	0.0122	0.0118
0 2 2	-----	0.00529	0.00432	0.00456	0.00478

Table 7.13: Survey of the Bragg reflections and integrated intensities (*IntegInt*) for the experimental (*Exp-100K*) and MD simulated (*ak2*) structures of cyclopentane in the low T range, between 0K and 35K, corresponding to phase III. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns. The integrated intensities of the peaks are normalised.

In figure 7.42 we see that the shape of the calculated patterns in the low temperatures range is very similar to that observed in the experimental pattern at 100K. Many positions of the peaks of the reflections [*hkl*] in the calculated patterns are slightly shifted with respect to the experiment, mainly towards lower angles with respect to the experimental positions. The amount of shift is between 0.2° and 0.6°. Some of the reflections are shifted by the same amount but towards higher values of 2θ with respect to the experimental ones ([*2 0 –2*]; [*-3 –1 2*]; [*4 0 –2*]). Finally some of the reflections in the calculated patterns fall at the same angle of the experimental ones ([*1 0 –2*]; [*2 0 0*]; [*2 1 –1*]; [*2 1 0*]; [*-2 –1 –3*]; [*-3 –1 –3*]). The intensities of the peaks (in table 7.13) match in reasonable agreement with the experiment.

As noted, a second different pattern is observed in the higher temperature range of the stable monoclinic phase III, from 40K up to 124K, as shown in figure 7.43. Peaks positions and integrated intensities are reported in table 7.14 and 7.15.

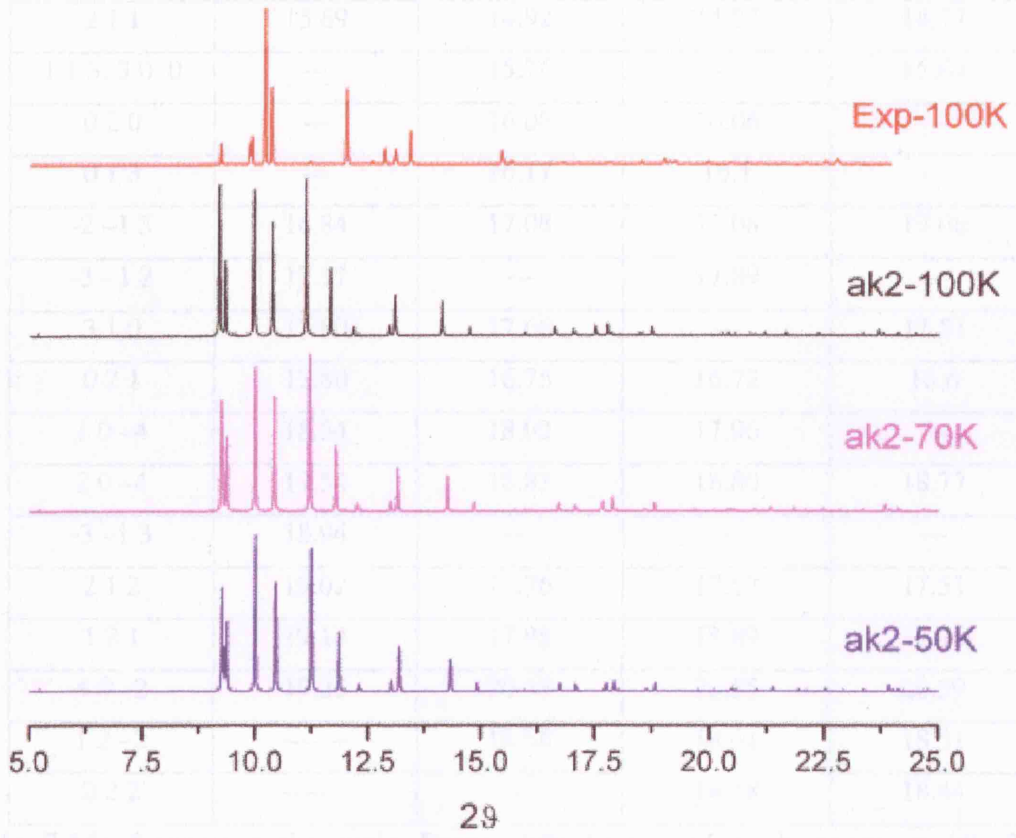


Fig. 7.43: Calculated XRD powder patterns of the MD simulated distorted phase III (*ak2*) in the higher temperature range, between 40K and 124K, compared to the experimental XRD powder pattern at 100K (*Exp* in red).

hkl	*2θ (Exp-100K)*	*2θ (ak2-50K)*	*2θ (ak2-70K)*	*2θ (ak2-100K)*
1 0 –2	9.24	9.38	9.25	9.36
0 1 1	9.88	9.28	9.36	9.20
0 0 2	9.93	9.32	9.29	9.20
1 1 –1	10.22	10.01	10	9.97
2 0 0	10.37	10.47	10.44	10.4
2 0 –2	11.23	---	15.73	12.03
1 1 1	12.03	11.26	11.22	11.13
-1 –1 2	12.60	---	12.28	12.29
2 1 -1	12.88	13.03	13.00	12.99
1 0 2	12.88	11.88	11.81	11.67
0 1 2	13.12	12.3	12.28	12.17
2 1 0	13.45	13.21	13.18	13.10
1 1 2	15.48	---	14.27	14.14
2 1 1	15.69	14.92	14.87	14.77
1 1 3; 3 0 0	---	15.76	---	15.64
0 2 0	---	16.08	16.06	---
0 1 3	---	16.17	16.1	---
-2 –1 3	16.84	17.08	17.08	17.06
-3 –1 2	17.17	---	17.89	---
3 1 0	17.80	17.68	---	17.51
0 2 1	17.80	16.75	16.72	16.6
1 0 –4	18.54	18.02	17.96	---
2 0 –4	19.54	18.83	18.80	18.77
-3 –1 3	18.94	---	---	---
2 1 2	19.02	17.76	17.67	17.51
1 2 1	19.14	17.93	17.89	17.76
4 0 –2	19.25	20.57	20.55	20.59
1 2 –2	-------	18.64	18.61	18.51
0 2 2	-------	---	18.58	18.44

Table 7.14: Survey of the main Bragg reflections and peak positions 2θ, for the experimental (*Exp-100K*) and MD simulated structures (*ak2*) of cyclopentane in the higher temperature range, between 40K and 124K, corresponding to distorted phase III. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns.

hkl	*IntegInt (Exp-100K)*	*IntegInt (ak2-50K)*	*IntegInt (ak2-70K)*	*IntegInt (ak2-100K)*
1 0 –2	0.0953	0.434	0.452	0.507
0 1 1	0.152	0.917	0.913	
0 0 2	0.154	0.321	0.452	0.974
1 1 –1	1	1	1	0.507
2 0 0	0.154	0.692	0.707	0.732
2 0 –2	0.0319	---	0.00692	0.0126
1 1 1	0.926	0.932	0.931	1
-1 –1 2	0.0438	---	0.0611	0.0148
2 1 -1	0.163	0.0692	0.0704	0.0737
1 0 2		0.368	0.386	0.436
0 1 2	---	0.0606	0.0611	0.0572
2 1 0	0.213	0.279	0.271	0.258
1 1 2	0.409	---	0.217	0.215
2 1 1	0.329	0.0578	0.0547	0.0537
1 1 3; 3 0 0	---	0.00820	0.00370	0.0142
0 2 0	---	0.00480	0.00496	---
0 1 3	---	0.00426	0.00565	---
-2 –1 3	0.0709	0.0450	0.0557	0.0576
-3 –1 2	0.0594	---	0.0778	---
3 1 0	0.0315	0.0148	---	0.0716
0 2 1	0.0522	0.0575	0.0557	0.0525
1 0 –4	---	0.0154	0.0155	---
2 0 –4	0.0737	0.0548	0.0543	0.0546
-3 –1 3	---	---	---	---
2 1 2	0.0144	0.0561	0.0703	0.0716
1 2 1	0.234	0.0816	0.0778	0.0576
4 0 –2	0.142	0.0141	0.0140	0.0201
1 2 –2	---	0.0274	0.0278	0.0152
0 2 2	---	---		0.0140

Table 7.15: Survey of the Bragg reflections and integrated intensities (*IntegInt*) for the experimental (*Exp-100K*) and MD simulated (*ak2*) structures of cyclopentane in the higher T range, between 40K and 124K, corresponding to distorted phase III. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns. The integrated intensities of the peaks are normalised.

On examining table 7.14, we see that the overall reproduction of the experimental pattern of phase III is less good. We note that in the higher temperature patterns we have almost all of the reflections contained in the experimental XRD pattern of the phase III. As in the low temperature patterns, we still have the majority of the reflections shifted towards lower angles compared to the corresponding experimental values. The shift in the high temperature pattern is larger than that observed in the low temperature pattern, being between 0.6° and 1.4°. In particular, we observe a systematic "swapping" of the peak positions relating to the [*1 0 –2*] and [*0 1 1*] with respect to the low temperature pattern. Again in the high temperature pattern there are some reflections shifted towards higher 2θ with respect to the experiment as observed in the low temperature range and other reflections at the same angular position with respect to the experiment. There are some experimental low intensity reflections at higher angles ($2\theta \geq 16^{\circ}$), which are completely missing in the high temperature pattern. A big discrepancy in the integrated intensities of the peaks is observable between the high temperature pattern and the experiment. The intensities of some reflections continue to be reasonably well reproduced like for those in the pattern corresponding to the low temperature range ([*2 0 –2*]; [*1 1 1*]; [*-1 –1 2*]; [*2 1 –2*]; [*1 0 2*]; [*2 1 0*]; [*1 1 2*]; [*-2 –1 3*]; [*-3 -1 2*]; [*3 1 0*]; [*0 2 1*]; [*2 0 –4*]), but some high temperature pattern reflections have a much larger intensities compared to the corresponding experimental data ([*1 0 –2*]; [*0 1 1*]; [*0 2 2*]; [*2 0 0*]; [*2 1 1*]; [*1 2 1*]; [*4 0 –2*]). Although still with high intensity, the [*1 1 –1*] reflection at 100K, has a lower integrated intensity in favour of [*1 1 1*].

The big difference between the intensity of the peaks of the higher temperature patterns, compared to the experimental XRD and calculated patterns in the low temperature range, is due to the presence of a strong thermal expansion, larger than that observed experimentally. The overestimation of the thermal expansion affects strongly the positions and the intensities of the peaks in the X-ray powder pattern, by causing a large shift of the peaks positions towards lower angles for specific reflections[41,42]. The increasing of the intensity with the thermal expansion is also related to the shift of the reflections towards low angles, which corresponds to a shift towards smaller reciprocal space vectors. In fact, closer to the reciprocal space origin (at smaller angles) the form factors of the atoms are larger and thus the calculated intensities increase[43]. As already discussed in the section *7.3.2*, the overestimation of the thermal expansion is mainly due to the parameters used in the repulsion-dispersion contribution to the potential, which

were fitted at room temperature. These imply an intrinsic temperature effect already contained in the potential, which makes the simulated system at high temperatures undergo a larger thermal expansion than the real system. Nevertheless, for the calculated XRD pattern of phase III in the range of high temperatures 40K-124K, it is important to note the presence of all the main experimental reflections. Thus these patterns still are representative of the ordered monoclinic phase III of cyclopentane, although under an excessive thermal expansion.

In the temperature range 125K - 127K, we calculated a different kind of pattern (figure 7.44) from that observed in the two ranges of temperatures related to the monoclinic, ordered phase III.

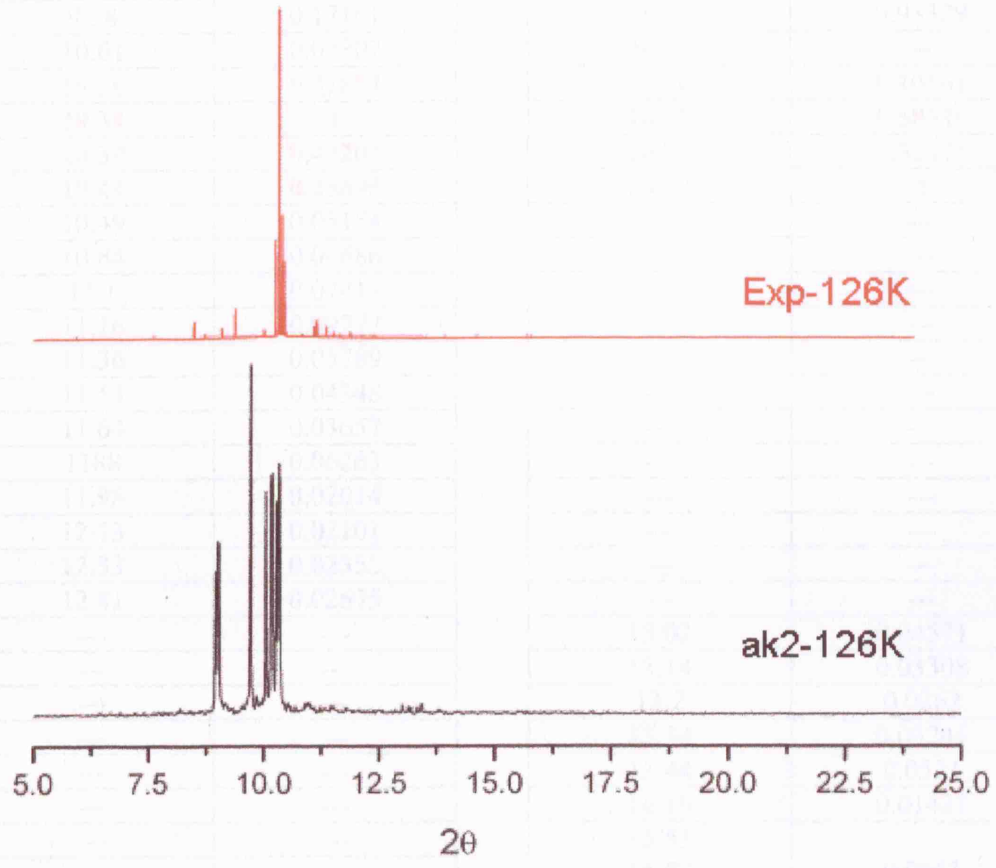


Fig. 7.44: Experimental (in red) and calculated (in black) XRD pattern at 126K, relating to the intermediate disordered phase II of cyclopentane.

The calculated pattern at 126K (figure 7.44 in black) displays significant background contribution, indicating the presence of disorder in the solid phase of cyclopentane at this temperature. This is also observed in the experimental powder pattern[2] (figure 7.44 in red). The two patterns show some similarities and some differences: two of the peaks at

low angles in the calculated pattern ($2\theta = 8.96^{o}$ and 9.02^{o}) have very high intensities compared to the low angles peaks in the experimental pattern. Furthermore, the experimental pattern has a larger number of low angles peaks, than the calculated one. All the experimental and calculated peaks are reported in table 7.16.

2θ (Exp126K)	*IntegInt (Exp126K)*		*2θ (ak2-126K)*	*IntegInt (ak2-126K)*
5.93	0.01141		---	---
5.97	0.01057		---	---
7.67	0.0228		7.21	0.01604
7.71	0.01065		7.4	0.01623
7.83	0.02047		---	---
8.49	0.07549		---	---
8.71	0.05367		8.96	0.36841
9.09	0.07795		9.015	0.73489
9.38	0.17161		9.72	0.93329
10.01	0.07807		9.85	--
10.26	**0.32894**		**10.06**	**0.10561**
10.34	**1**		**10.17**	**0.58526**
10.39	**0.47201**		**10.29**	**0.52474**
10.44	**0.25695**		**10.33**	**1**
10.49	0.05154		---	---
10.84	0.04686		---	---
11.1	0.07417		---	---
11.16	0.09777		---	---
11.36	0.05789		---	---
11.53	0.04348		---	---
11.64	0.03657		---	---
1188	0.06263		---	---
11.98	0.02014		---	---
12.13	0.02101		---	---
12.33	0.02355		---	---
12.41	0.02675		---	---
---	---		13.02	0.04571
---	---		13.14	0.03308
---	---		13.2	0.0262
---	---		13.34	0.03241
---	---		13.44	0.0531
---	---		14.16	0.01421
---	---		15.53	
---	---		15.57	0.0463
---	---		15.61	
---	---		15.7	0.01409

Table 7.16: Peak positions and relative integrated intensities (*IntegInt*) for the experimental and the calculated XRD powder pattern at 126K of cyclopentane solid phase II. Each box of the table corresponds to a peak or overlapped peaks in the XRD powder patterns. The values in red are relative to the common peaks between the calculated and experimental powder pattern. The integrated intensities of the peaks are normalised.

On the other hand, the common characteristic of the two patterns are the four peaks, centred around the position $2\theta \sim 10.2^{o}$-10.3^{o} (highlighted in red colour in table 7.16 and shown in figure 7.45), which are very close to each other and are present only in the powder pattern of the phase II. The integrated intensities of these four calculated peaks match well with the experimental integrated intensities. At higher angles, on the right part of the experimental pattern there is another dense group of peaks from $2\theta \sim 13.02^{o}$ to $2\theta \sim 13.44^{o}$, with the integrated intensities of these peaks being much smaller than the intensities of the peaks around 10.3^{o}. The same group was observed also in the experimental XRD pattern, although in a different angular interval, from $2\theta \sim 11.1^{o}$ to $2\theta \sim 12.41^{o}$, with the experimental integrated intensities being of the same order of the intensities of the calculated peaks in interval 13.02^{o} -13.44^{o}. The latter group of peaks (shown for the two patterns in figure 7.46) could be associated to the same group reflections in the two patterns, with the theoretical reflections shifted in the calculated pattern with respect to the experimental ones of c.a. 2^{o}. The reason of the shift might be the limitations in the intermolecular potential, especially at high temperatures, as already noted for the monoclinic ordered phase III.

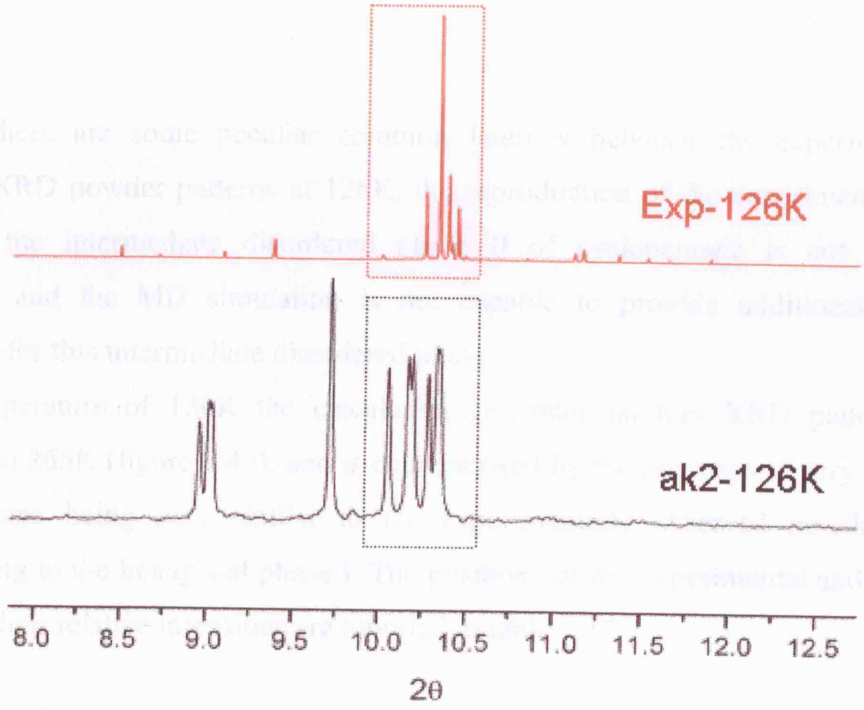


Fig. 7.45: Enlargements between $8.0^{o} < 2\theta < 12.0^{o}$ of the experimental and calculated XRD patterns of cyclopentane corresponding to the intermediate plastic phase II at 126K. The highlighted regions correspond to the two equivalent groupings of reflections for the two patterns, at $10.0^{o} < 2\theta < 10.5^{o}$.

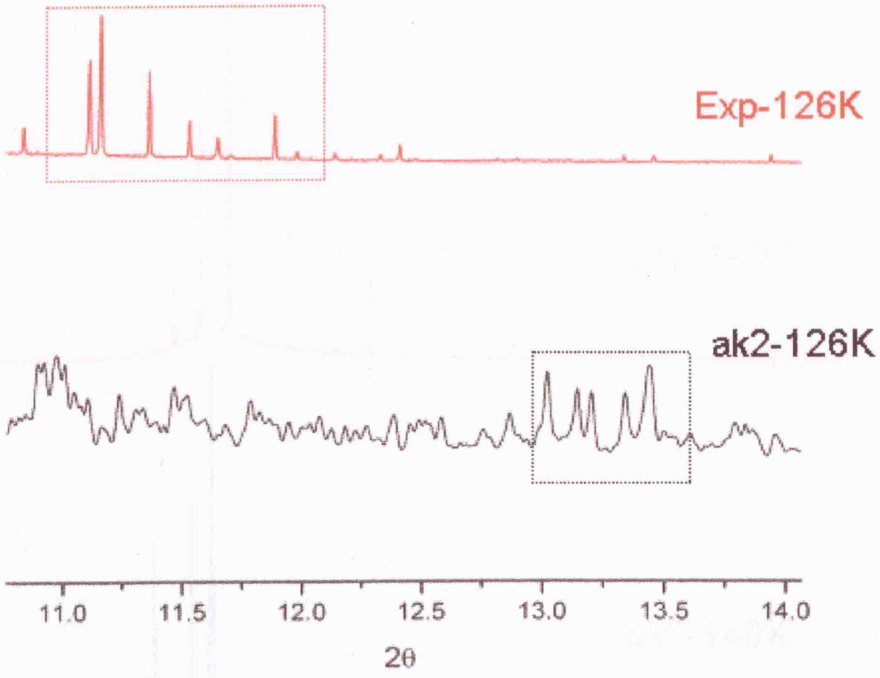


Fig. 7.46: Enlargements between $11.0^{o} < 2\theta < 14.0^{o}$ of the experimental and calculated XRD patterns of cyclopentane corresponding to the intermediate plastic phase II at 126K. The highlighted regions correspond to hypothetical equivalent groupings of reflections for the two patterns, respectively in the range 11.10^{o}-11.98^{o} (Exp-126K) and 13.02^{o}-13.44^{o} (calculated ak2-126K).

Although there are some peculiar common features between the experimental and calculated XRD powder patterns at 126K, the reproduction of the experimental powder pattern of the intermediate disordered phase II of cyclopentane is not completely satisfactory and the MD simulation is not capable to provide additional structural information for this intermediate disordered phase.

At the temperature of 130K the calculation generates another XRD pattern, which persists up to 265K (figure 7.47), and is characterised by the presence of very few peaks, with its shape being very similar to the experimentally observed powder pattern, corresponding to the hexagonal phase I. The positions of the experimental and calculated peaks with their relative intensities are reported in table 7.17.

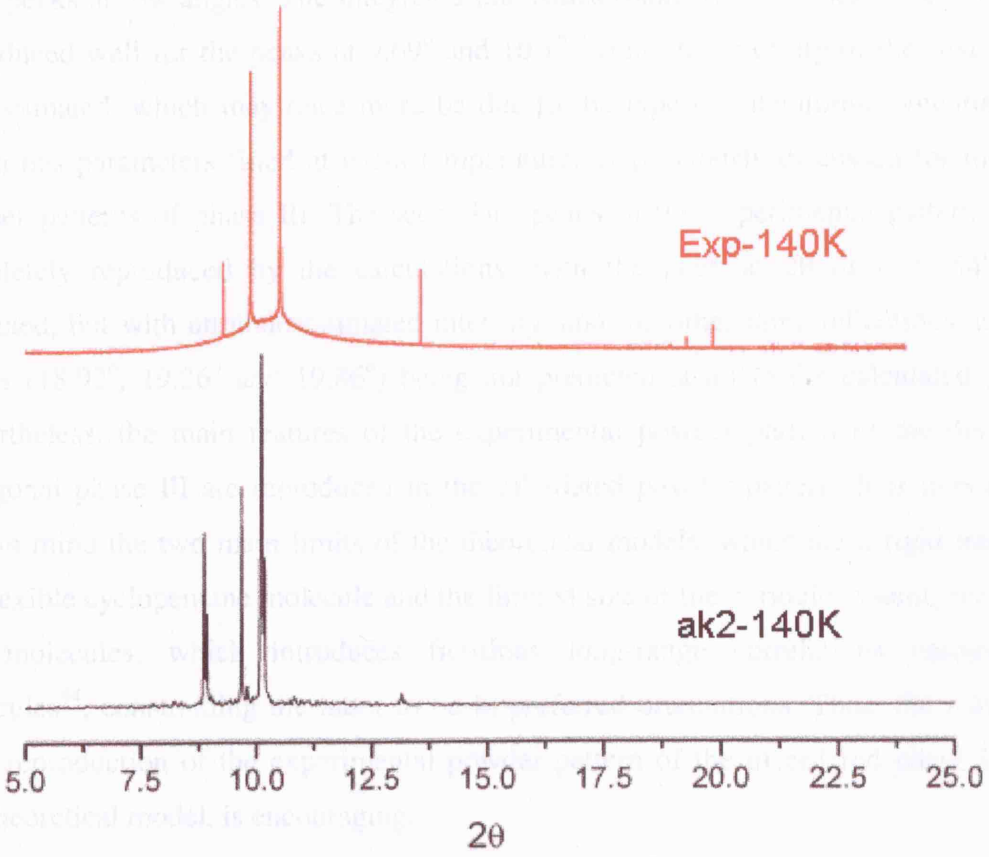


Fig. 7.47: Experimental XRD powder pattern of hexagonal disordered phase I (in red) and calculated pattern at 140K (in black).

2θ (Exp-140K)	*IntegInt (Exp-140K)*		*2θ (ak2-140K)*	*IntegInt (ak2-140K)*
9.27	0.08415		8.88	0.37564
9.87	0.60606		8.92	
10.51	1		9.69	0.2881
13.57	0.11464		10.1	1
18.92	0.02797		10.18	
19.26	0.08141		13.14	0.04211
19.83	0.03301		18.45	0.00587

Table 7.17: peak positions and integrated intensities (*IntegInt*) in the experimental and calculated XRD powder pattern of cyclopentane disordered phase I. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns. The integrated intensities of the peaks are normalised.

As in the calculated powder pattern of phase II at 126K, the calculated pattern at 160K displays significant background contribution, characteristic of a highly disordered system, corresponding to the high temperature disordered phase I^2. We see from the comparison of the two XRD powder patterns in figure 7.47 that the calculated powder pattern at 140K reproduces experiment reasonably well, especially concerning the three

main peaks at low angles. The integrated intensities (table 7.17) of these three peaks are reproduced well for the peaks at 9.69° and 10.1°, while the intensity of the first peak is overestimated, which may once more be due to the type of interatomic potential used, which has parameters fitted at room temperature, as previously discussed for the XRD powder patterns of phase III. The secondary peaks in the experimental pattern are not completely reproduced by the calculations, with the peak at 2θ of ~ 13.54° being predicted, but with an underestimated intensity, and the other three reflections at higher angles (18.92°, 19.26° and 19.86°) being not predicted at all in the calculated pattern. Nevertheless, the main features of the experimental powder pattern of the disordered hexagonal phase III are reproduced in the calculated powder pattern. It is important to bear in mind the two main limits of the theoretical models, which are a rigid model for the flexible cyclopentane molecule and the limited size of the periodic system, containing 180 molecules, which introduces fictitious long-range correlations between the molecules[44], constraining the latter to be in preferred orientations. Thus, the reasonably good reproduction of the experimental powder pattern of the disordered phase I within this theoretical model, is encouraging.

Finally we report the calculated XRD powder patterns for the higher temperatures (from 200K to 268K), at which we observed a super-heated disordered phase I and their comparison with the calculated pattern of phase I at 160K, in order to see if there is an effect of the super-heating on the XRD pattern of the system. Table 7.18 reports all the peak positions of the reflections of phase I at different temperatures and in table 7.19 the corresponding integrated intensities, while the corresponding powder patterns are illustrated in figure 7.48.

h k l	*2θ (ak2-160K)*	*2θ (ak2-200K)*	*2θ (ak2-230K)*	*2θ (ak2-260K)*	*2θ (ak2-265K)*	*2θ (ak2-268K)*
0 0 2	8.92	8.82	8.77	8.68	8.66	8.65
0 1 1	8.88	8.89	8.72	8.68	8.62	8.66
0 1 -1	8.88	8.92	8.72	8.64	8.62	8.75
2 0 0	9.69	9.39	9.42	9.32	9.32	9.35
1 0 -2	10.14	9.98	9.96	9.84	9.83	9.73
1 0 2	10.10	10.01	9.96	9.80	9.83	9.94
1 1 -1	10.10	10.09	9.94	9.84	9.81	9.82
1 1 1	10.10	10.09	9.99	9.83	9.81	9.84
1 -1 -1	10.10	10.09	9.91	9.83	9.81	9.86
1 -1 1	10.10	10.09	9.87	9.80	9.81	10.04
2 0 -2	13.14	12.9	12.89	12.68	12.72	12.60
2 0 2	---	12.9	12.89	12.68	12.72	12.92
2 1 -1	13.14	13.00	12.82	12.68	12.72	12.66
2 1 1	13.17	13.00	12.82	12.68	12.72	12.76
2 -1 -1	13.14	13.00	12.82	12.68	12.72	12.79
2 -1 1	13.14	13.00	12.82	12.68	12.72	13.00
0 1 3	---	15.33	15.18	---	---	
0 1 -3	---	---	15.25	---	---	14.98
2 -1 2	---	---	---	---	---	15.12
0 2 0	---	15.52	---	---	---	15.16
0 2 1	---	---	---	---	---	15.72
1 1 3	---	---	---	---	---	15.75
1 2 0	---	---	---	---	---	15.79
3 1 0	---	---	---	---	---	15.87
1 -2 0	---	---	---	---	---	15.90
1 2 -1	---	---	16.51	---	---	---
1 2 2	18.45	---	---	---	---	---

Table 7.18: Survey of the main Bragg reflections and peak positions 2θ, for the MD simulated phase I, and super-heated phase I of cyclopentane from 160K to 268K. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns.

h k l	*IntegInt (ak2-160K)*	*IntegInt (ak2-200K)*	*IntegInt (ak2-230K)*	*IntegInt (ak2-260K)*	*IntegInt (ak2-265K)*	*IntegInt (ak2-268K)*
0 0 2	0.376	0.374	0.365	0.408	0.352	0.481
0 1 1						
0 1 -1						0.344
2 0 0	0.288	0.293	0.342	0.400	0.300	0.641
1 0 -2	1.00	1.00	1.00	1.00	1.00	0.375
1 0 2						0.355
1 1 -1						1.000
1 1 1						
1 -1 -1						
1 -1 1						0.327
2 0 -2	0.0421	0.0483	0.0332	0.0348	0.0352	0.0632
2 0 2						0.0236
2 1 -1			0.0140			0.0632
2 1 1						0.0297
2 -1 -1						
2 -1 1						0.0318
0 1 3	---	0.0171	0.00568	---	---	0.0122
0 1 -3	---			---	---	
2 -1 2	---	---	---	---	---	0.0214
0 2 0	---	0.00417	---	---	---	
0 2 1	---	---	---	---	---	0.0220
1 1 3	---	---	---	---	---	
1 2 0	---	---	---	---	---	
3 1 0	---	---	---	---	---	0.0261
1 -2 0	---	---	---	---	---	
1 2 -1	---	---	0.00777	---	---	---
1 2 2	0.00527	---	---	---	---	---

Table 7.19: Survey of the Bragg reflections and integrated intensities (*IntegInt*) for the MD simulated phase I, and super-heated phase I of cyclopentane from 160K to 268K. Each box of the table corresponds to a peak or a group of overlapped peaks in the XRD powder patterns. The integrated intensities of the peaks are normalised.

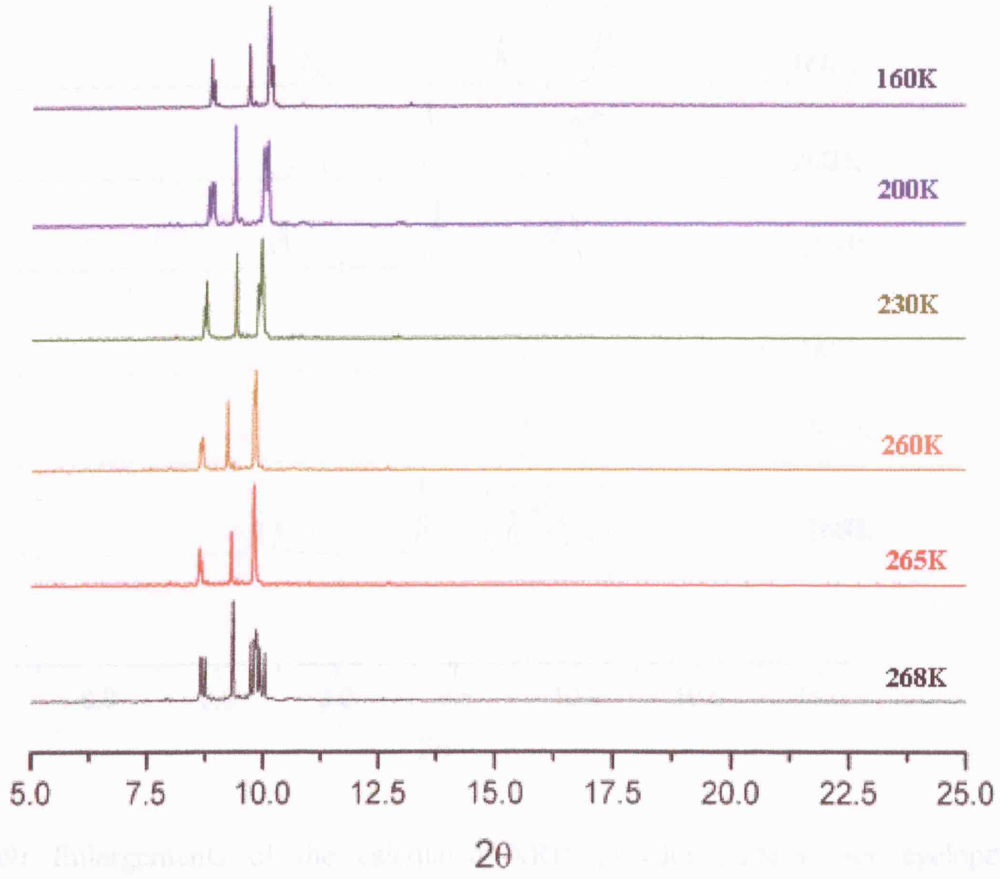


Fig. 7.48: Calculated XRD powder patterns for the super-heated hexagonal disordered phase I, and the super-heated phase I at high temperatures.

From the data in the tables 7.18, 7.19 and from the figure 7.48, we note that the main peaks of the hexagonal disordered phase I are still present in the super-heated phase at temperatures of 200K, 230K, 260K and 265K. The peaks of the main reflections shift progressively to smaller values of 2θ, owing to the thermal expansion of the crystal (as shown in the plot of the volumes *vs* temperature in figure 7.18). The integrated intensities of these peaks are more or less constant with increasing the temperature and they are very close to the intensities observed in the pattern at 160K, whilst a progressive slight increment of the noise with temperature is observable in the baseline of the patterns. At temperature of 268K the calculated powder pattern appears a little different compared to the calculated pattern from 160K up to 265K, with the difference mainly concerning the peak around $2\theta \sim 10.0^{o}$ (figure 7.48: the peak highlighted by the dotted square for the pattern at 268K) as clearly results in the enlargement of the patterns shown in figure 7.49.

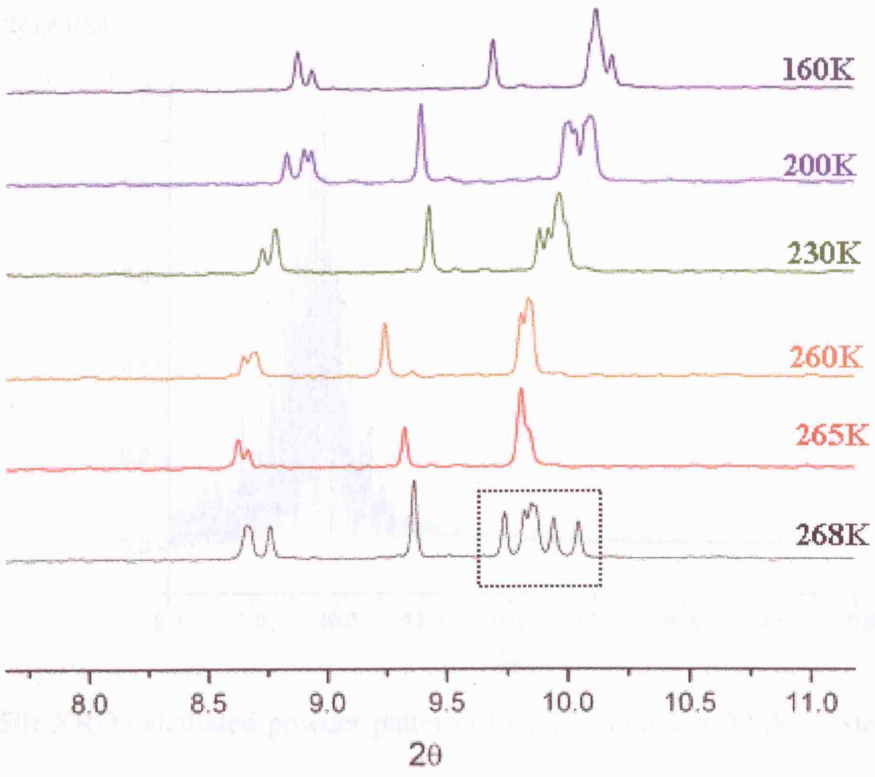


Fig.7.49: Enlargements of the calculated XRD powder patterns for cyclopentane hexagonal disordered phase I at 160K and in the super-heated state between 180K and 268K. The dotted square highlights the peak of the XRD pattern at 268K, which substantially differs from the corresponding peaks in the patterns at the other temperatures.

At 268K, we observe a splitting of some peaks (also evident in table 7.18 and 7.19), in particular for those at $2\theta \sim 8.6^{o}$, $2\theta \sim 9.9^{o}$, which correspond respectively to the groups of reflections [*0 0 2; 0 1 1; 0 1 –1*] and [*1 0 –2; 1 0 2; 1 1 –1; 1 1 1; 1 –1 –1; 1 –1*]. This is probably due to a break of the symmetry of the system, which changes from the hexagonal symmetry of the phase I to a system with lower symmetry, when approaches the liquid phase, occurring at 270K. This pre-melting stage, with loss of symmetry by the system is in accordance with an homogeneous nucleation mechanism of melting with the formation of a pre-nucleus of liquid particles, generated by strong elastic lattice instabilities inside the crystal[31].

At 270K the powder pattern (figure 7.50) has extensive noise without any structural information and with the presence of a very broaden large peak, which is similar to that observed in amorphous phases[45,46] or liquid phases with complete lack of long-range

order, but presence of short-range order. The symmetry of the crystalline state is completely lost.

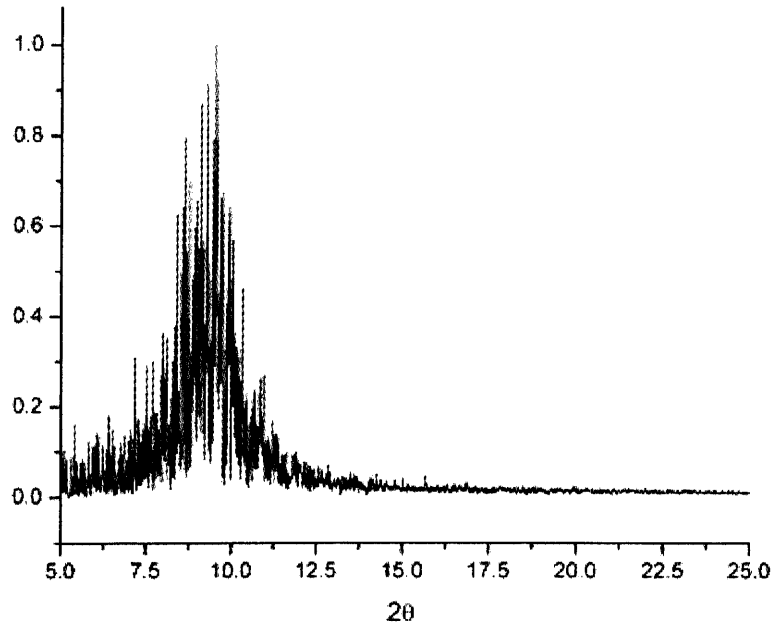


Fig. 7.50: XRD calculated powder pattern of cyclopentane at 270K: system in the liquid phase.

To summarise, in accordance with experimental data, the calculation of the diffraction patterns in the range of temperatures between 100K and 170K gives three different types, reflecting the presence of three different phases of solid cyclopentane. The change in the shape of the theoretical spectra with temperature is coherent with the change observed in the structural parameters with temperature. The X-Ray diffraction pattern for the calculated distorted monoclinic ordered phase III is observed up to 124K. The pattern for the intermediate disordered phase II is only observed at 125K and 126K, whilst from 127K the pattern evolves into that typically observed for the hexagonal disordered phase I, which persists up to 265K. This trend moves away from experiment, in which the XRD powder pattern relative to the phase II is observed in a wider range of temperatures, between 120K and 132K. Hence the MD simulated system switches from monoclinic to hexagonal phase more rapidly with temperature and this is consistent with the trend of the calculated lattice parameters with temperature.

The calculated X-Ray diffraction pattern for the MD simulated system for monoclinic ordered phase III differs from experiment at the corresponding temperature (figure 7.43). As discussed in the section *7.3.3*, the calculated phase III at 100K has a distorted monoclinic unit cell, compared to the experimental one at the same temperature, which

explains the difference between the observed and the calculated XRD patterns. However the calculated XRD patterns at $T \leq 35K$ reproduce reasonably well the corresponding experimental data (figure 7.42). A good approximation is similarly obtained for the calculated powder pattern of hexagonal disorderd phase I (figure 7.47), especially for the main high intensities reflections, whose peaks positions and integrated intensities are in good agreement with experiment. The calculated powder pattern also shows the quality of the reproduction of the hexagonal symmetry of the high temperature orientationally disordered phase I, by the MD simulated heating process. This pattern is also observed in the calculated super-heated phase I, up to 265K, with small shifts of the peak posistions towards smaller angles, while at 268K, the calculated pattern shows a strong splitting of the peaks, which can be associated with a symmetry reduction close to the melting point. This latter is observed in the MD calculations at 270K and the XRD calculated pattern of the system is consistent with a completely structurally disordered phase, such a liquid.

Overall the X-Ray diffraction data indicate the adequate reproduction by the simulations of the monoclinic ordered phase III and hexagonal disordered phase I of cyclopentane and of the transition between these two phases on heating the system.

The experimental X-Ray diffraction pattern at 126K associated with the intermediate phase II is partially reproduced by the MD simulation. In fact the simulated pattern at 126K shows some particular characteristics of the experimental pattern, but also some discrepancies. The main deficiency consists of a poorer distribution of peaks in the calculated pattern, which makes it closer to the characteristic pattern of the hexagonal phase I. Considering also the very restrict range of temperatures in which the simulated phase II occurs, it is likely that the simulation is able to reproduced only partially the intermediate disordered phase II. The latter can be interpreted as a sort of intermediate metastable state approaching the disorderd phase I from the ordered monoclinic phase, which implies a modification in the symmetry of the structure, which cannot be represented by either a monoclinic or a hexagonal cell. Consequently the quality of reproduction of the intermediate phase is very sensitive to the approximations involved in the simulation, such as the rigid molecule approximation and the limited size of the simulation cell., which may prevent a more complete observation of the proper symmetry of phase II, resulting only in a partially adeguate reproduction of the experimentally observed X-Ray powder pattern.

7.3.7 Orientational dynamics of the molecules at different temperatures

The study of the rotational dynamics of the molecules in the structures at different temperatures, from a qualitative point of view, can be very helpful in characterising and classifying the different phases of cyclopentane, especially understanding the differences between the two disordered plastic phases, phase II and phase I. A useful tool for a qualitative analysis of the dynamics of the molecules at different temperatures is the visualisation of the trajectory file obtained from the MD simulations. From these animations we can recognise two main different rotational motions of the molecules associated to two different ranges of temperature. The first is observed in the range corresponding to the monoclinic ordered phase III, consists of molecules librating around their centres of mass, with highly correlated orientations. The second is observed starting from 125K, temperature at which the *order-disorder* phase transition occurs, and it consists of molecules, which have more free rotational movement around their centres of mass and characterised by rotations of various wide angular degrees and by a variety of different orientations. These molecular rotations seem to be not continuous and not isotropically distributed around all the possible orientations. For an individual molecule we can observe a large rotation, followed by a librational movement, which keeps the molecule along the new orientation for a definite amount of time, before to change again with a new large re-orientation. Although a very useful and direct tool to understand the dynamics of the molecules in the crystals at different temperature, the visual animation is not publishable in a conventional format and it does not give an easy understanding of the collective behaviour of the molecules in the system, in their rotational motion. Hence it is necessary to analyse the atomic positions in the trajectory file with a different method in order to obtain more quantitative and detailed information about the rotational dynamics of the molecules of the system, at different temperatures.

In examining the rotational molecular motion, an analysis of the properties of the system must be made on data arising from simulations in the NVE ensemble, in order to avoid the fluctuations of the cell parameters in NST simulations, which could affect the positions of the atoms in the system, from which the information about the rotational dynamics has to be derived. The NVE simulations were performed, choosing the starting configuration with the same method adopted for the X-ray powder pattern calculations, and using the same interatomic potential, the same timestep of 0.001ps, but a longer equilibration time (up to 150ps) and longer production (up to 50ps), for a total number of 500 configurations of the system in the NVE trajectory file.

The first method, for characterising the dynamical molecular motion of the system, is discussed in section 7.3.7a, and it consists of a statistical analysis of the rotations of the 180 molecules in the supercell as function of temperature. This analysis can help us to understand the nature of the dynamics of the molecules in the system at different temperatures, revealing the presence of partial disorder, in which modality the latter is generated. The second method is discussed in section 7.3.7b and it consists of calculating the orientational distribution function (ODF) of the system at different temperatures from the atomic positions in the trajectory file of the simulated system. This method can reveal preferential orientations in the rotational motion of the molecules, describing in greater details the orientational disorder of the system at different temperatures.

7.3.7.a Statistical Analysis of the molecular rotations.

We defined for each molecule in the supercell two internal vectors, connecting two not adjacent carbon atoms of the molecular ring, which are C1-C4 and C2-C5, and the respective associated vectors are denoted as r_{12} and r_{25}, (figure 7.51: blue and red vectors).

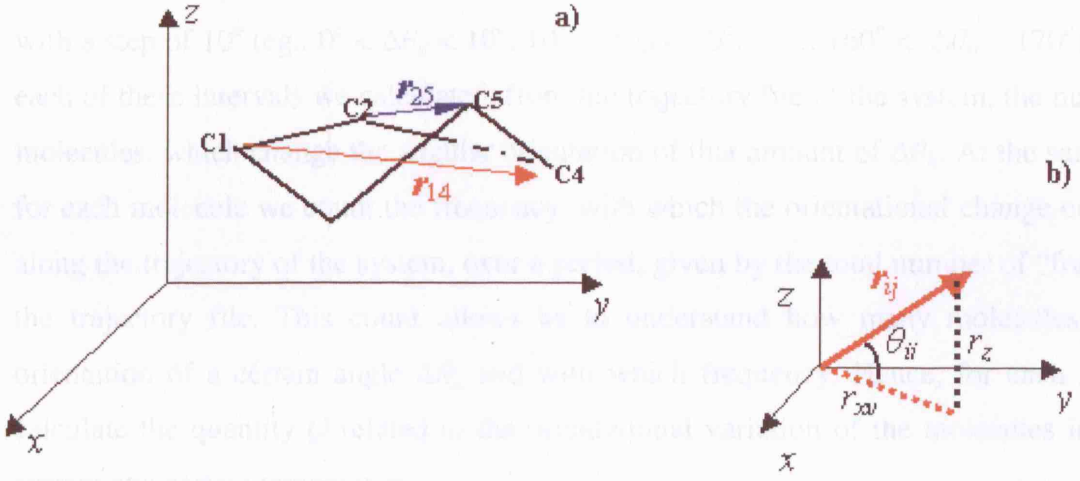


Fig. 7.51: a) Schematic representation of the internal molecular vectors considered in the model to study the orientation of the molecules in the MD simulated crystals at different **b)** schematic illustration of the calculation of the orientational angles θ_{ij} of the cyclopentane molecule. The indexes i and j are related to the couple of atoms in the molecule, considered in the calculation of the intra-molecular vector r_{ij}.

Furthermore, we consider the angles between these two vectors and the $(x;y)$ plane, defined by the external cartesian reference system as illustrated in figure 7.51a, which are respectively denoted θ_{14} and θ_{25}, and define all the possible orientations of the molecule

for a complete description of the rotational dynamics of each molecule in the space, with respect to the cartesian plane (x; y). These angles were calculated through trigonometric formula:

$$\theta_{ij} = \arctan\frac{r_z}{r_{xy}} \quad , \qquad [7.5]$$

where i and j represent the generic couples of indices relative to the carbon atoms belonging to the molecular ring(1-4 and 2-5).

For each of the 180 molecules in the system, we calculate the variation in the orientation of the molecule as:

$$\Delta\theta_{ij}(t) = \theta_{ij}(t) - \theta_{ij}^{0} \quad . \qquad [7.6]$$

This is the difference between the θ_{ij} in the initial configuration in the trajectory of the system (θ_{ij}^{0}) and the value of the angle $\theta_{ij}(t)$ at the time t. In this way the initial orientation for all molecules results 0° with respect to the (x;y) plane in the box of the MD simulation.

From the values of these angular differences $\Delta\theta_{ij}$, we performed two kinds of analysis. First we carried out a statistical analysis about the possible orientational variations $\Delta\theta_{ij}(t)$ at different temperatures. For each temperature we consider different intervals of $\Delta\theta_{ij}(t)$, with a step of 10° (eg.: $0° < \Delta\theta_{ij} < 10°$; $10° < \Delta\theta_{ij} < 20°$;.....$160° < \Delta\theta_{ij} < 170°$) and for each of these intervals we calculated, from the trajectory file of the system, the number of molecules, which change the angular orientation of that amount of $\Delta\theta_{ij}$. At the same time, for each molecule we count the frequency, with which the orientational change occurs all along the trajectory of the system, over a period, given by the total number of "frames" in the trajectory file. This count allows us to understand how many molecules change orientation of a certain angle $\Delta\theta_{ij}$ and with which frequency. Hence, for each $\Delta\theta_{ij}$, we calculate the quantity Q related to the orientational variation of the molecules in all the system at a certain temperature:

$$Q = \frac{NumMolecul\ es(\Delta\theta_{ij})}{TotalNumOf\ Molecules} \cdot \frac{NumFrames\ (\Delta\theta_{ij})}{TotalNumOf\ Frames} = \frac{NumMolecul\ es(\Delta\theta_{ij})}{180} \cdot \frac{NumFrames\ (\Delta\theta_{ij})}{200} \qquad [7.7]$$

We plot these data using histograms at different temperatures for specific values of angular intervals. The data relative to the histograms (figures 7.52-7.54) are reported in table7.20.

angles	*100K*	*120K*	*122K*	*124K*	*125K*	*126K*	*127K*	*128K*	*132K*	*134K*	*136K*	*140K*	*150K*	*160K*	*170K*	*200K*	*230K*	*260K*	*265K*	*268K*
$0 < \Delta\theta < 10$	149.43	141.78	135.41	134.21	66.15	57.27	57.68	62.11	59.9	52.70	47.92	53.22	50.61	42.67	41.07	35.70	38.20	34.04	35.64	35.96
$10 < \Delta\theta < 20$	26.34	34.34	37.74	37.58	37.92	37.64	34.20	36.53	35.42	34.23	33.70	35.46	31.36	28.46	29.82	25.18	25.55	23.61	25.54	24.90
$20 < \Delta\theta < 30$	1.16	2.34	4.137	3.66	23.98	24.23	25.32	25.23	22.94	24.87	24.83	24.12	24.62	24.50	23.36	22.00	23.34	22.16	23.07	22.21
$30 < \Delta\theta < 40$	0.030	8.9E-3	0.092	0.25	14.08	15.06	16.55	14.27	16.89	18.67	18.72	17.98	18.54	20.87	20.18	20.29	19.09	20.24	19.26	20.32
$40 < \Delta\theta < 50$			1.2E-3	7.1E-3	7.41	9.78	11.84	9.58	11.63	11.72	13.48	12.36	13.66	16.37	18.19	17.54	17.42	16.60	17.28	17.56
$50 < \Delta\theta < 60$			2.5E-4	9.1E-3	4.28	7.43	6.67	6.56	6.85	7.45	9.42	7.49	10.31	11.85	13.25	15.13	14.42	14.22	15.14	15.33
$60 < \Delta\theta < 70$				1.2E-3	2.83	3.93	3.62	4.079	4.13	4.59	5.19	4.37	6.80	7.60	9.56	12.40	10.86	11.98	12.47	12.06
$70 < \Delta\theta < 80$				2.5E-4	1.32	1.75	1.94	2.35	2.88	2.90	3.39	2.87	3.52	5.84	5.76	9.58	7.676	10.79	9.19	9.36
$80 < \Delta\theta < 90$				8.3E-5	0.56	0.65	0.95	0.50	1.24	1.49	1.60	1.70	2.23	3.19	2.99	5.52	5.84	7.70	5.90	6.30
$90 < \Delta\theta < 100$					0.24	0.24	0.45	0.12	0.35	0.70	0.57	0.86	0.65	1.30	1.28	2.83	2.89	3.96	3.39	3.52
$100 < \Delta\theta < 110$					0.058	0.10	0.17	0.024	0.15	0.23	0.36	0.40	0.26	0.55	0.60	1.66	1.52	2.11	1.94	1.78
$110 < \Delta\theta < 120$					0.048	0.10	0.062	3.0E-3	0.046	0.12	0.14	0.12	0.099	0.35	0.17	0.66	0.74	1.28	0.98	0.96
$120 < \Delta\theta < 130$						0.053	0.029	2.3E-3	0.012	0.053	0.029	0.058	0.028	0.098	0.062	0.22	0.45	0.60	0.38	0.38
$130 < \Delta\theta < 140$						0.012	0.045	5.6E-3	8.0E-3	0.022	5.7E-3	0.014	0.019	0.015	0.036	0.031	0.13	0.20	0.15	0.13
$140 < \Delta\theta < 150$						4.4E-3	6.5E-3	5.0E-4	5.6E-5	5.0E-4	4.0E-3	6.0E-3	0.013	6.0E-3	4.9E-3	0.021	0.030	0.076	0.033	0.052
$150 < \Delta\theta < 160$						8.9E-4	2.5E-3	2.2E-4			3.3E-4	4.4E-4	9.6E-3	5.6E-5	3.8E-4	6.0E-3	9.5E-3	7.5E-3	0.010	4.2E-3
$160 < \Delta\theta < 170$								1.1E-4			5.6E-5	3.1E-4	5.0E-4		1.5E-3	9.4E-4	1.4E-3	5.0E-4	4.4E-4	4.2E-4
$170 < \Delta\theta < 180$																	2.8E-5	2.8E-5		

Table 7.20: Q values for the statistics of the angular orientations distribution of the molecules in the cyclopentane supercell 3x5x3 at different temperatures.

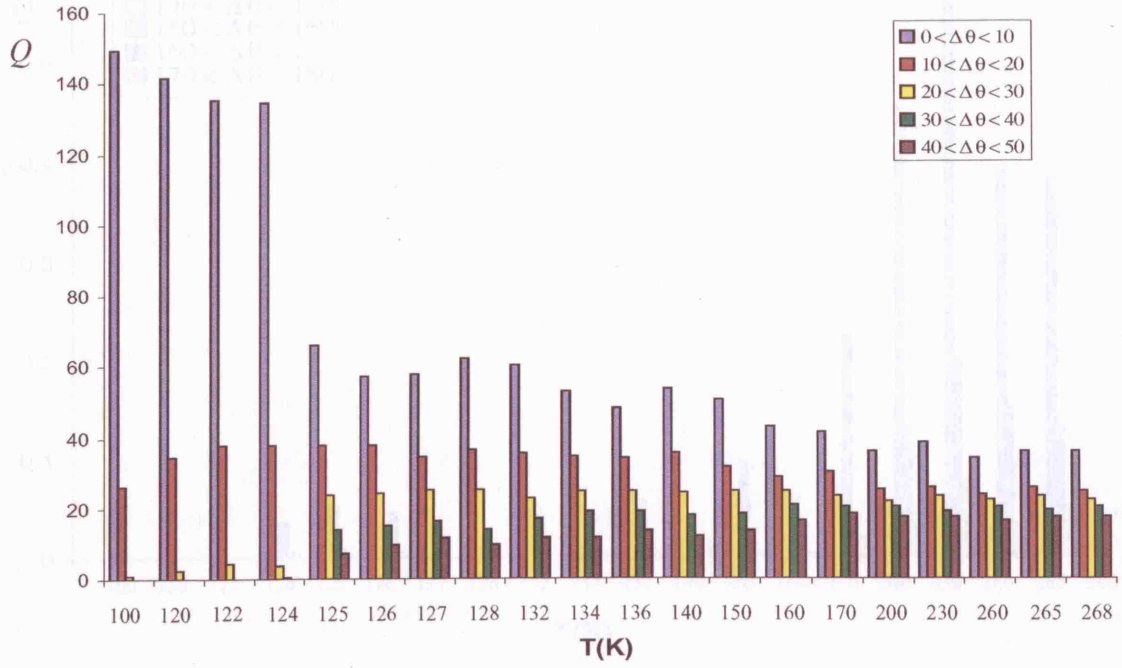


Fig. 7.52: Histogram of the molecular rotational angle distribution in the cyclopentane crystals at different temperatures: low rotational angles $\Delta\theta_{25}$.

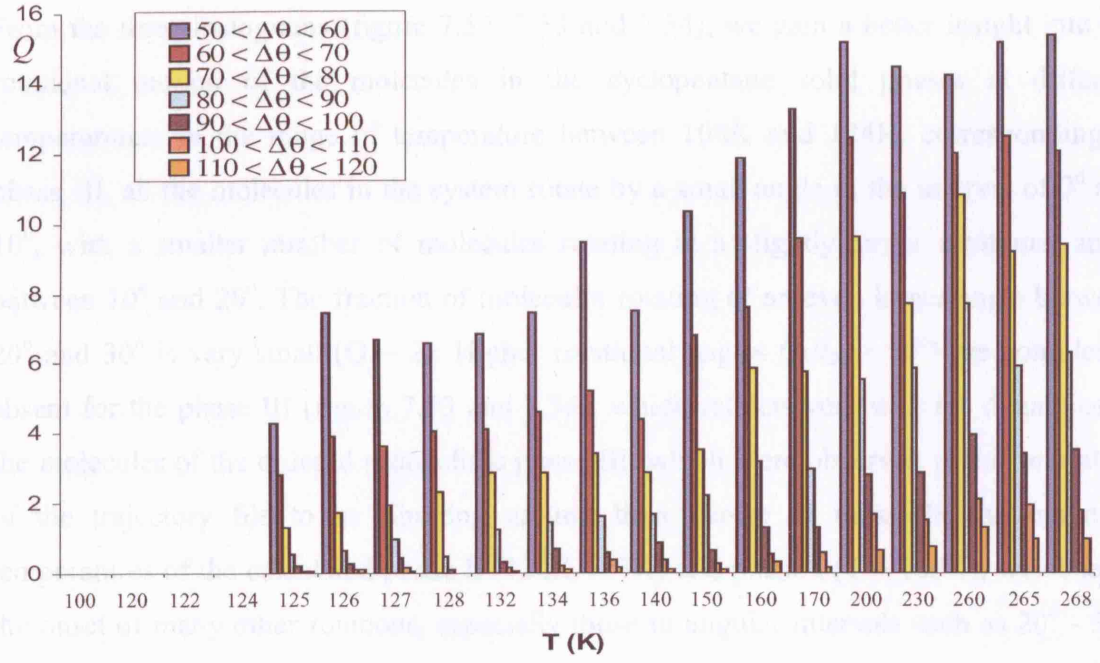


Fig.7.53: Histogram of the molecular rotational angle distribution in the cyclopentane crystals at different temperatures: Higher rotational angles $\Delta\theta_{25}$.

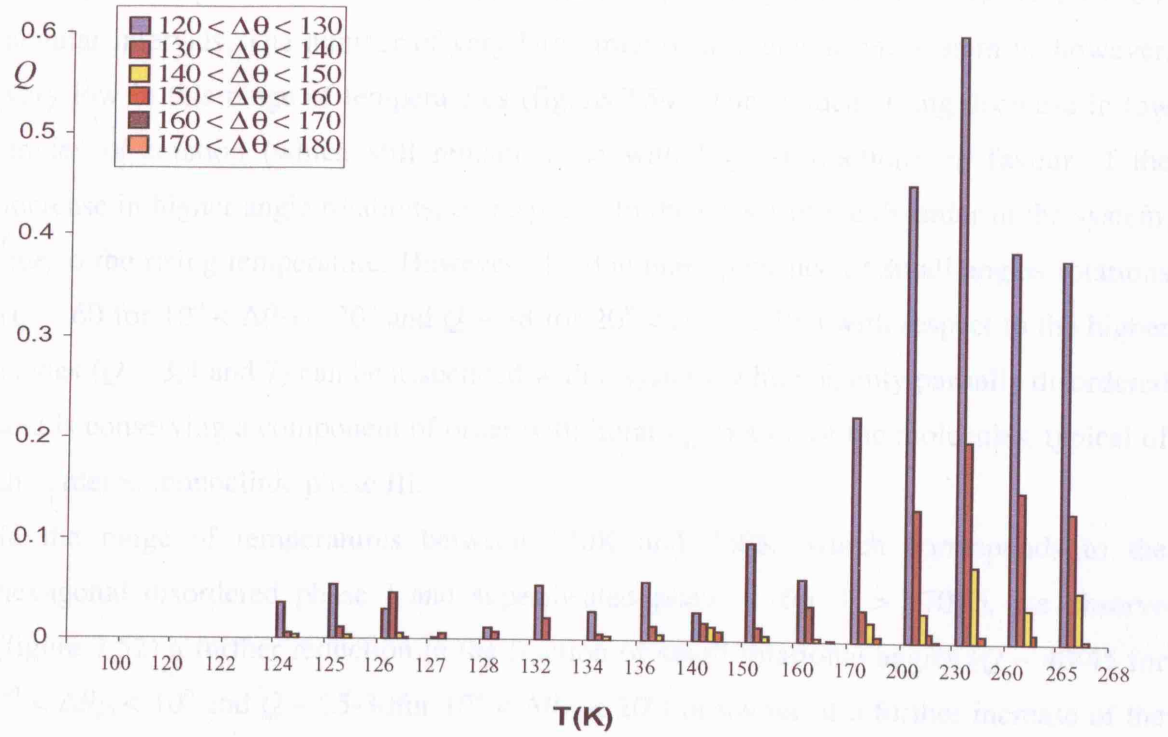


Fig.7.54: Histogram of the molecular rotational angle distribution in the cyclopentane crystals at different temperatures: very large rotational angles $\Delta\theta_{25}$.

From the three histograms (figure 7.52, 7.53 and 7.54), we gain a better insight into the rotational motion of the molecules in the cyclopentane solid phases at different temperatures. In the range of temperature between 100K and 124K, corresponding to phase III, all the molecules in the system rotate by a small angle in the interval of 0° and 10°, with a smaller number of molecules rotating at a slightly larger rotational angle between 10° and 20°. The fraction of molecules rotating of an even larger angle between 20° and 30° is very small (Q ~ 2). Higher rotational angles ($\Delta\theta_{25} > 50°$) are completely absent for the phase III (figure 7.53 and 7.54), which reflects very well the dynamics of the molecules of the ordered monoclinic phase III, which were observed in the animation of the trajectory file to be librating around their centre of mass. In the range of temperatures of the calculated phase II (126K-127K) and phase I (T > 128K), we observe the onset of many other rotations, especially those in angular intervals such as 20° - 30°, 30° - 40°, 40° - 50°. In particular we see (figure 7.52) a strong reduction of the small angle rotations (0° - 10°), in favour principally of the rotations of 20°-30° and 30°-40°, whilst the fraction of rotations in the angular interval 10°-20° remains almost constant with temperature. We also see (figure 7.53) the on-set of the higher rotational angles (50°

$< \Delta\theta_{25} < 80^{o}$), with values of $Q \sim 3$, 4 and 7 respectively for 50^{o}-60^{o}, 60^{o}-70^{o}, 70^{o}-80^{o} angular intervals. The number of very large rotational angles in the system is, however, very low in this range of temperatures (figure 7.54). The sudden strong decrease in low angles of rotation (which still remain those with highest fraction), in favour of the increase in higher angle rotations, corresponds to the on-set of the disorder in the system, due to the rising temperature. However, the dominant presence of small angles rotations ($Q \sim 60$ for $10^{o} < \Delta\theta_{25} < 20^{o}$ and $Q \sim 38$ for $20^{o} < \Delta\theta_{25} < 30^{o}$) with respect to the higher angles ($Q \sim 3$,4 and 7) can be associated with a system, which is only partially disordered and is conserving a component of order with librating motion of the molecules, typical of the ordered monoclinic phase III.

In the range of temperatures between 140K and 268K, which corresponds to the hexagonal disordered phase I and super-heated phase I (for T > 170K), we observe (figure 7.52) a further reduction in the fraction of small rotational angles ($Q \sim 40$-45 for $0^{o} < \Delta\theta_{25} < 10^{o}$ and $Q \sim 25$-30for $10^{o} < \Delta\theta_{25} < 20^{o}$) in favour of a further increase of the higher angles of rotation (figure 7.53), up to $Q \sim 16$ for the angular interval $10^{o} < \Delta\theta_{25} < 20^{o}$ in the superheated phase I. The fraction of all higher rotational angles (in the interval $60^{o} < \Delta\theta_{25} < 90^{o}$) increases with temperature (in figure 7.53). From the histogram in figure 7.54 we see that very high angle rotations ($120^{o} < \Delta\theta_{25} < 140^{o}$) also occur, starting from 160K and more strongly in the superheated phase, although their fractions ($Q \sim 0.4$–0.6) remain much smaller than those related to the rotations in the interval $50^{o} < \Delta\theta_{25} < 90^{o}$. As in the previous ranges of temperatures, we can also still observe a consistent fraction of small angle molecular rotations. Nevertheless, especially in the superheated phase I, the strong increment in the fraction of the high angle rotations, which is close to the fraction of the small angle rotations ($Q_{small\text{-}angle} \sim 25$-30; $Q_{high\text{-}angles} \sim 15$-16) can be interpreted as in increase of orientational disorder in the system because of the temperature effect. Finally we note that the trends observed in the three histograms relating to $\Delta\theta_{25}$ is the same for the three corresponding to $\Delta\theta_{14}$.

The second analysis, which we performed to investigate the rotational dynamics of the molecules in the cyclopentane solid phases, consists of the calculation of the average rotational angles $\Delta\theta_{25}$ and $\Delta\theta_{14}$ over the trajectory of the system, for each molecule. The average is performed over all of the frames of the trajectory of the system, resulting from an NVE simulation for 50ps, with a time-step of 0.001ps, after an equilibration time of 100ps. The time interval of 0.1ps is used in printing different configurations in the trajectory file, for a total number of 500 configurations. To take into account the different

frequency, at which the intervals of the angular variations $\Delta\theta_{ij}$ are covered, we performed a weighted average, with a weight s, consisting of the frequency, which a certain rotational angle $\Delta\theta_{ij}$ occurs with, in angular intervals from 0° to 180°, selected with a step of 5 ° (eg. $0° < \Delta\theta_{ij} < 5°$; $5° < \Delta\theta_{ij} < 10°$; $10° < \Delta\theta_{ij} < 20°$.... $175° < \Delta\theta_{ij} < 180°$). This average is mathematically written as:

$$\left\langle \Delta\theta_{ij} \right\rangle_s = \frac{\sum_{F=1}^{500} \Delta\theta_{ij}{}^F \cdot s^F}{\sum_{F=1}^{500} s^F} \quad , \qquad [7.8]$$

with s^F given by the expression:

$$s^F = \frac{N^F(\Delta\theta_{ij})}{N_{Frames}} = \frac{N^F(\Delta\theta_{ij})}{500} \quad , \qquad [7.9]$$

where $N^F(\Delta\theta_{ij})$ is the number of times the molecule performs a rotation of an angle $\Delta\theta_{ij}$ in the interval $5n < \Delta\theta_{ij} < 5(n+1)$, with $n = 0, 1, 2, ...36$.

As a result of this calculation, we plot the values of the weighted average orientations of all the molecules at the temperatures, corresponding to the three different phases (figures 7.55-7.58) and we use graphs ($\Delta\theta_{14}$, $\Delta\theta_{25}$) reported in figures 7.55-7.58, in which each point corresponds to a molecule, with that particular average rotational "state".

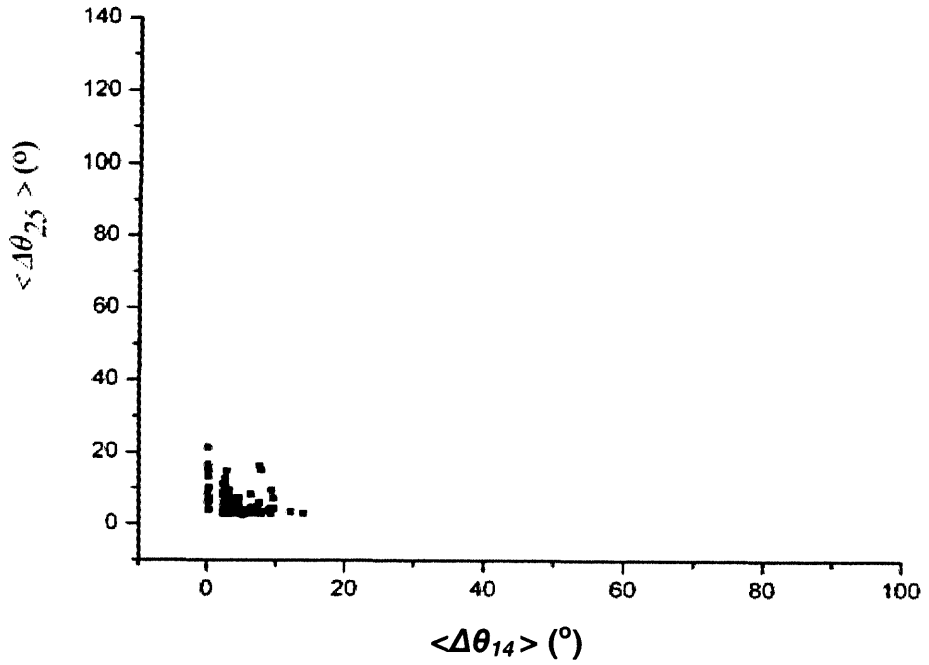


Fig. 7.55: Plots of the average rotations of the molecules in the solid monoclinic ordered phase III of cyclopentane, at 100K.

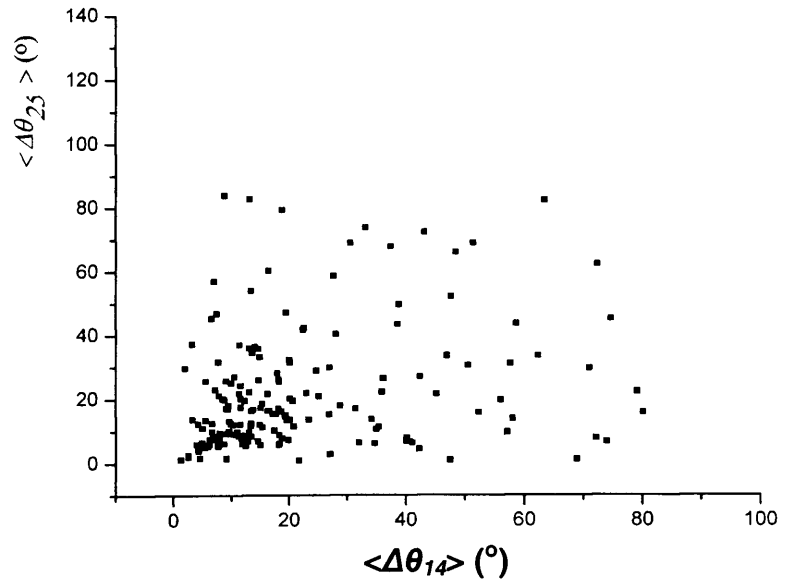


Fig. 7.56: Plots of the average rotations of the molecules in the intermediate plastic phase II of cyclopentane, at 125K.

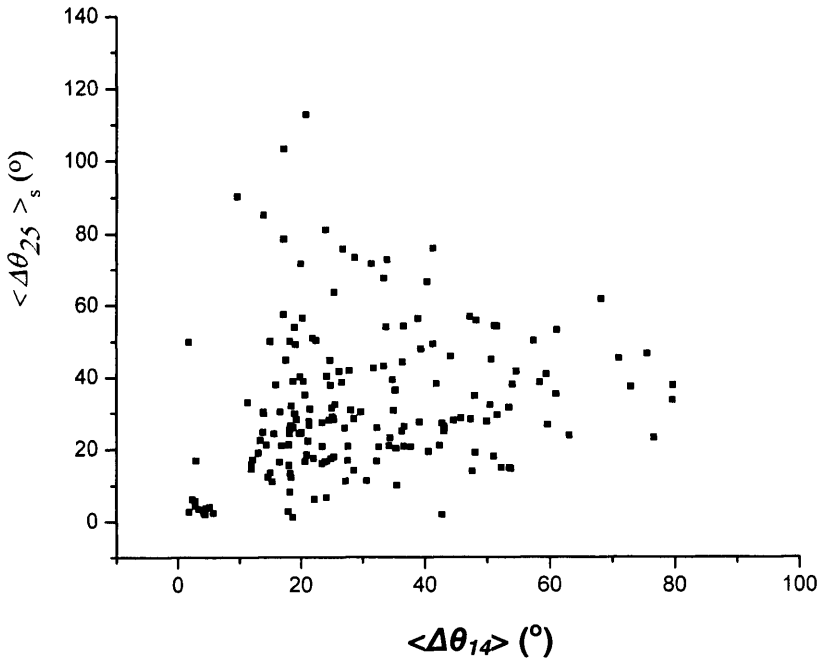


Fig. 7.57: Plots of the average rotations of the molecules in the hexagonal plastic phase I of cyclopentane, at 160K.

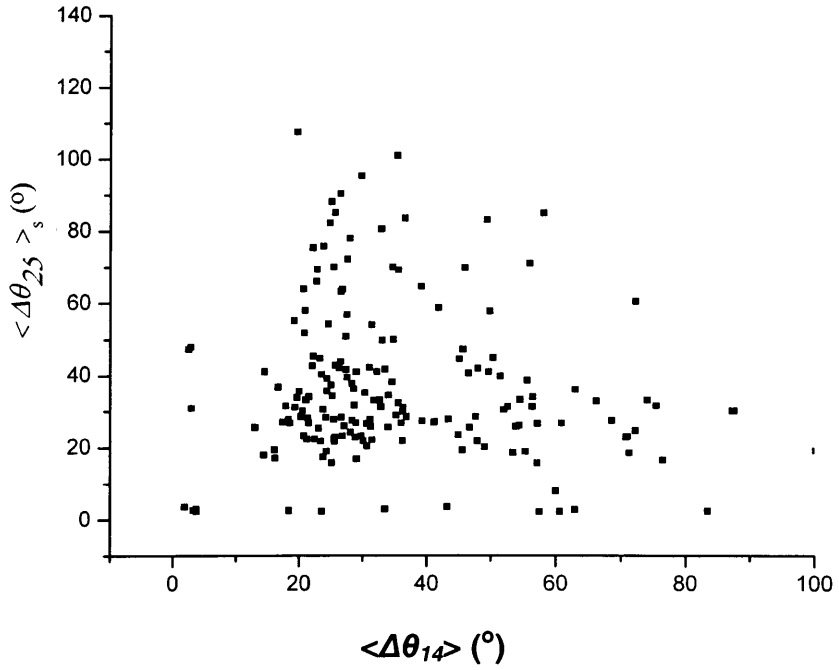


Fig. 7.58: Plots of the average rotations of the molecules in the super-heated hexagonal plastic phase I of cyclopentane, at 230K.

Figure 7.55 shows that all the molecules in the system at 100K are in a rotational state at low angle, which is characterised by very small rotational angles, mainly around 5°-10°, with few rotations of 20° and 15° respectively in the θ_{14} and θ_{25} angle. This behaviour reflects a high level of order in the system with all of the molecules changing their orientation only by a few degrees. In figure 7.56 there is a clear change in the "rotational "state of the molecules in the system, with some molecules undergoing large orientations in one preferred orientation (θ_{14} or θ_{25}) up to 80°. However most of the molecules are still in the small-angle rotational state, with variations in one of the angles mainly around 10°-20°. This plot represents a partially disordered system with only a fraction of the molecules suffering large variations in their spatial orientation which can be associated with the intermediate, partially disordered phase II, which is experimentally not well characterised. The last two plots in figure 7.57 and 7.58 show the presence of a larger number of molecules, undergoing large average rotations. Many points in the previous plots (7.55 and 7.56) have moved towards larger angles in $\Delta\theta_{14}$ and $\Delta\theta_{25}$. This result implies an increasing number of molecules with large average rotations, mainly concentrated around 30°- 40°-50°, but with some average rotations up to 100°-110°,

especially for θ_{25} angle, while the average rotations of θ_{14} angle are not larger than 80°. An asymmetry appears in the increase of the orientations of the molecules in the disorder system at high temperature (T > 160K), which suggests that there are few preferred orientations even in the high temperature disordered system, phase I.

7.3.7.b Orientational Distribution Function (ODF) calculation.

To give an insight into the rotational dynamics of the molecules in the crystal structures of cyclopentane at different temperatures, we calculated the orientational distribution function (ODF) of the molecules and observed the way, in which it varies with changing temperature. In the following section we give a definition of the ODF and show the kind of information we can extract from this function and its variation with temperature.

The ODF is a function of polar coordinates $Z(\theta; \varphi)$ and it reveals the most probable directions of specified molecular vectors in space (figure 7.59).

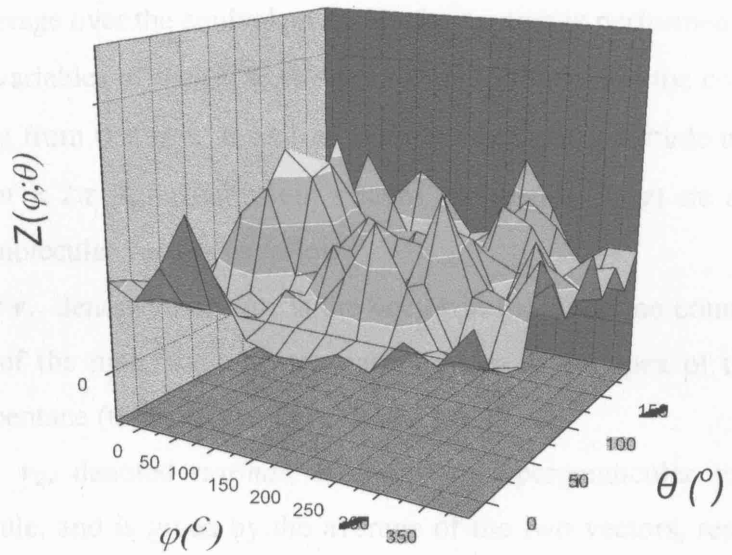


Fig. 7.59: Example of ODF $Z(\theta; \varphi)$ $Z(\theta; \varphi)$, for the plastic phase II of cyclopentane at 125K. The colours correspond to different values of ODF, which are associated with different values of orientational probability.

This function can be expressed as linear expansion of functions with spherical symmetry. In many studies, a set of *Gaussian* orthonormal functions was used in expanding the ODF[47], but an expansion in terms of *real spherical harmonic* functions $Y_l^m(\theta;\varphi)$ is also very common[48,49]. These are eigen-functions of the angular momentum operator L^2

associated to the rotations[50] and are angular portions of an orthogonal set of functions, which are solutions to Laplace's equation[51,52], in a system described by spherical coordinates *(r; θ; φ)*, having the general expression:

$$Y_l^m(\theta,\varphi) = \sqrt{\frac{(2l+1)}{4\pi}\cdot\frac{(l-m)!}{(l+m)!}}\cdot e^{im\varphi}\cdot P_l^m(cos\theta) \quad , \qquad [7.10]$$

where the functions $P_l^m(cos\theta)$ are *associated Legendre polynomials* [51]. The program we used for calculating the ODF, written by Dr. M. Leslie, calculates the ODF function $Z(\theta; \varphi)$ as linear expansion of *spherical harmonics*:

$$Z(\theta;\varphi) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} C_{lm}\cdot Y_l^m(\theta;\varphi) \quad . \qquad [7.11]$$

Following the standard approach[48], the program truncates the expansion at l = 12. The coefficients C_{lm} are calculated as ensemble averages of the corresponding spherical harmonics for a certain molecular vector:

$$C_{lm} = \left\langle Y_l^m \right\rangle \quad . \qquad [7.12]$$

Finally an average over the equivalent molecular vectors is performed.

The angular variables of the ODF are defined as follows: θ is the colatitude or the polar angle, ranging from $0 < \theta < \pi$, and φ is the azimuth or longitude angle, ranging in the interval $0 < \varphi < 2\pi$. Referring to our system, the angles $(\theta; \varphi)$ are defined with respect two internal molecular vectors as follows:

a) vector v_1, denoted *director*, is the vector in the direction connecting the centre of mass of the molecule with the carbon atom at the apex of the envelope-shaped cyclopentane (the blue vector in figure 7.60).

b) vector v_2, denoted *normal*, is the vector perpendicular to the plane of the molecule, and is given by the average of the two vectors, respectively normal to the two planes, which define the smallest torsion angle (figure 7.60).

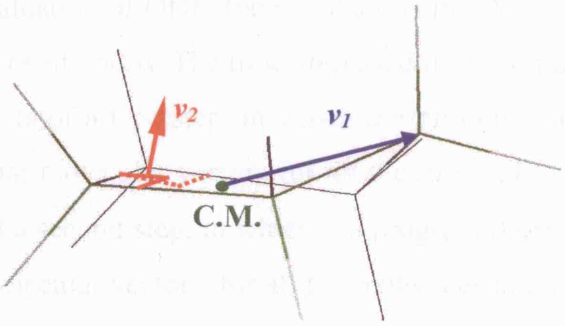


Fig.7.60: The two intermolecular vectors in the cyclopentane molecule, considered in the calculation of the ODF: *director* vector v_1 (blu), and the *normal* vector v_2 (black).

We define θ as the angle between one molecular vector and the ***b*** axis and φ the angle between the projection of the molecular vector on the ***ac*** plane and the ***c*** axis (figure 7.61).

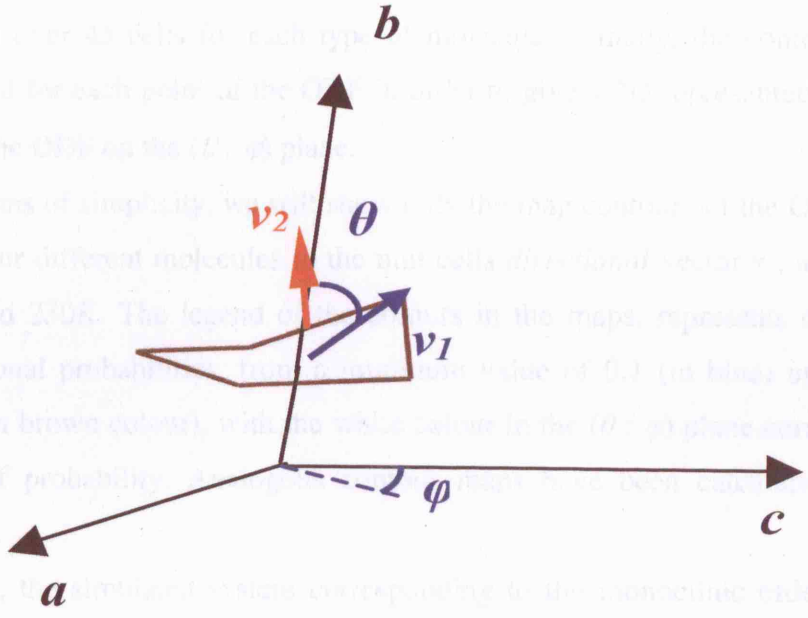


Fig. 7.61: The two angles (θ; φ) defined for the cyclopentane in calculating the ODF as spherical harmonics expansion. The angles θ and φ relative to the molecular vector v_1 (in blue) are shown in the figure. Similar angles are define for the vector v_2 (in red).

The angles θ, φ are defined in the same way either for the director or the normal internal molecular vectors. The procedure used by the program extracts all the information to calculate the ODF from the trajectory file obtained by the simulations of cyclopentane in

the NVE ensemble, already described for the calculation of the X-ray powder patterns. In the case of the evaluation of ODF, the simulations in NVE are extended up to 50ps, after an equilibration time of 150ps. The time-step used in the simulation is 0.001ps.

The method consists of a first step, in which the program calculates the centres of mass and the two internal molecular vectors for all the molecules and for all the frames in the trajectory file, and a second step, in which the program determines the angles θ and φ for each of the two molecular vectors, for all the molecules and for all the 500 configurations in the trajectory of the system. The ODF functions $Z(\theta; \varphi)$ are successively computed for each pair of angles $(\theta; \varphi)$ and then the average of the ODF over the 500 configurations is calculated. Hence at this stage the program calculates an ODF, averaged over 500 configurations of the system, for each of the 180 molecules in the 3x5x3 supercell. The 180 molecules in the supercell are classifiable into four unique types, which correspond to the four unique molecules in the original unit cell of the crystal structure. The 3x5x3 supercell consists of 45 unit cells. Therefore in the next stage the calculated ODF is averaged over 45 cells for each type of molecules. Finally, the contour of the ODF is calculated for each point of the ODF in order to give a 2D representation as a projection map of the ODF on the $(\theta; \varphi)$ plane.

For reasons of simplicity, we will show only the map contours of the ODF relative to one of the four different molecules in the unit cells *directional* vector v_1, at 100K, 125K and 160K and 230K. The legend of the colours in the maps, represents different values of orientational probabilities, from a minimum value of 0.1 (in blue) up to values bigger than 1 (in brown colour), with the white colour in the $(\theta; \varphi)$ plane corresponding to zero values of probability. Analogous contour maps have been calculated for the *normal vector* v_2.

At 100K, the simulated system corresponding to the monoclinic ordered phase III, the contour map (figure 7.62), relating to one of the four unique molecules in the unit cell of cyclopentane phase III, corresponds to a unique peak of orientational probability in a restricted range of values of $(\theta; \varphi)$.

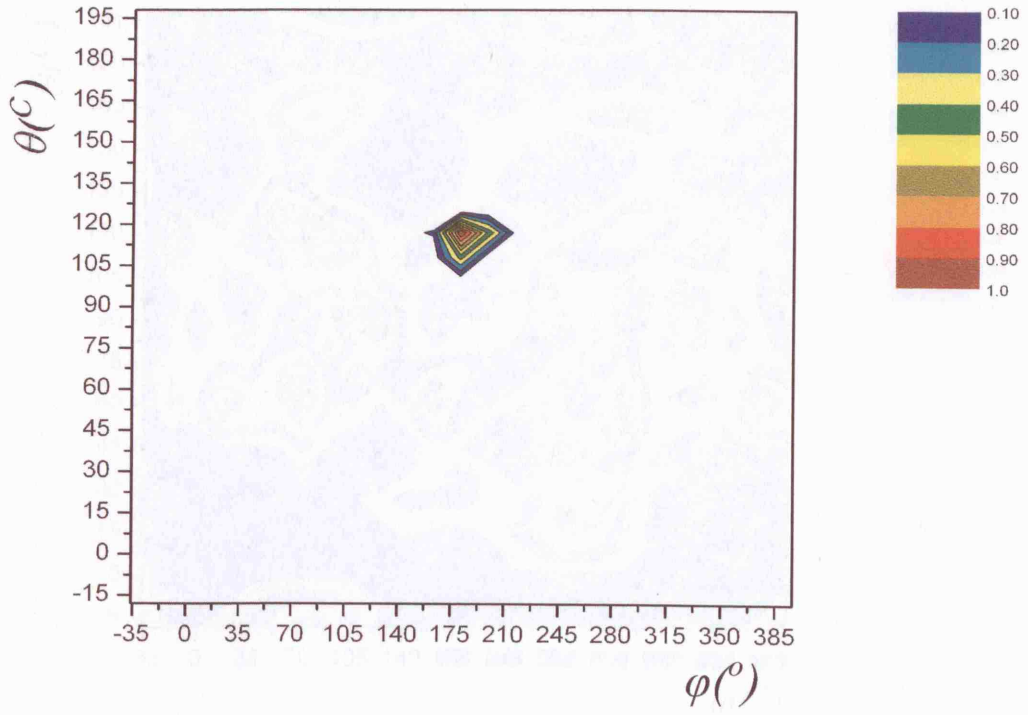


Fig. 7.62: Contour map of the ODF of the *directional* vector v_1 for one of the four unique molecules in the unit cell of the monoclinic ordered phase III of cyclopentane at 100K.

We note that the *directional vector* points in space with a unique specific preferential orientation ($\theta \sim 120^{o}$; $\varphi \sim 180^{o}$) with respect to the *b* axis. All the four types of molecules in the monoclinic unit cell are oriented along a preferential direction, which does not substantially change in time and such an ODF is typical for ordered systems, in which the thermal motion does not change the orientation of the molecules.

At 125K, in correspondence of phase II, we observe a completely different scenario: the contour map of the ODF (figure 7.63) becomes much more complicated.

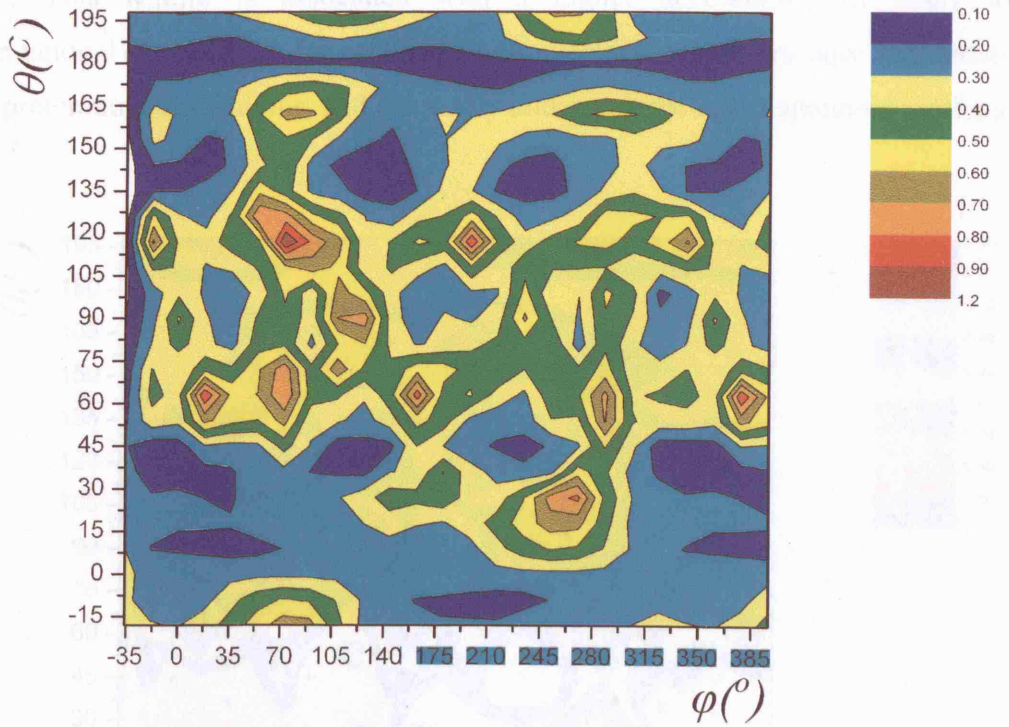


Fig. 7.63: Contour map of the ODF of the *directional* vector v_1 of a molecule in the intermediate plastic phase II of cyclopentane at 125K.

There are still preferred orientations, showed by the presence of localised peaks at specific angles in the ODF map (red-brown spots in figure 7.63), but there are more than at 100K. Considering the θ angle between the *directional* vector and the ***b*** axis for $\theta \sim 120^{\circ}$, we observe three values of φ ($\sim 75^{\circ}$; $\sim 205^{\circ}$ and $\sim 340^{\circ}$) with a high probability. Other preferred orientations are observed, as indicated by four peaks at $\theta \sim 60^{\circ}$ ($\varphi \sim 20^{\circ}$, 70°, 150°, 280°) and two other possible preferred orientations are apparent in the peaks at ($\theta \sim 90^{\circ}$; $\varphi \sim 110^{\circ}$) and ($\theta \sim 30^{\circ}$; $\varphi \sim 270$). Moreover we note a larger distribution, associated with a wide *non*–zero probability range of orientations for the molecules. The orientational probability for these areas is small (blue, light-blue and green areas in the contour map). This kind of plot indicates a system, in which preferential orientations in the molecular rotations dominate, among a very high number of limitedly accessible orientations.

The contour map of the ODF at 160K (figure 7.64), which corresponds to the plastic phase I, is similar to the one at 126K, and is characterised by a variety of peaks in the ODF, which corresponds to high preferred orientations. Nevertheless we note an important increase in the low angle orientations, as shown by the change in the colour

from blue (figure 7.63) to green and yellow (figure 7.64) for a large area of the contour map. This feature is associated with a higher accessibility of many low-angle orientational states (θ; φ) for cyclopentane molecules, which are now less constrained in the preferential orientations, and can easily undergo large re-orientational motions.

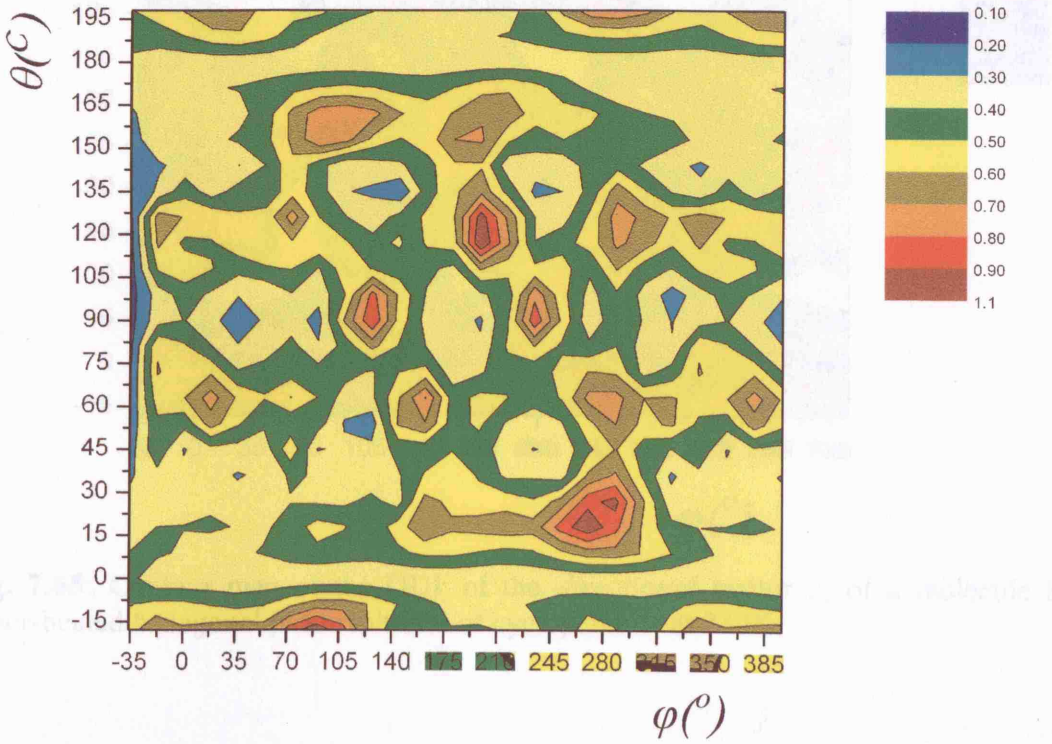


Fig. 7.64: Contour map of the ODF of the *directional* vector v_1 of a molecule in the hexagonal plastic phase I of cyclopentane at 160K.

Finally, the contour map of the super-heated phase I at high temperature (figure 9: 230K), shows an enlargement of the peaks of preferred orientations (brown and red colour regions in the map), corresponding to a wider molecular rotational motion around its centre of mass.

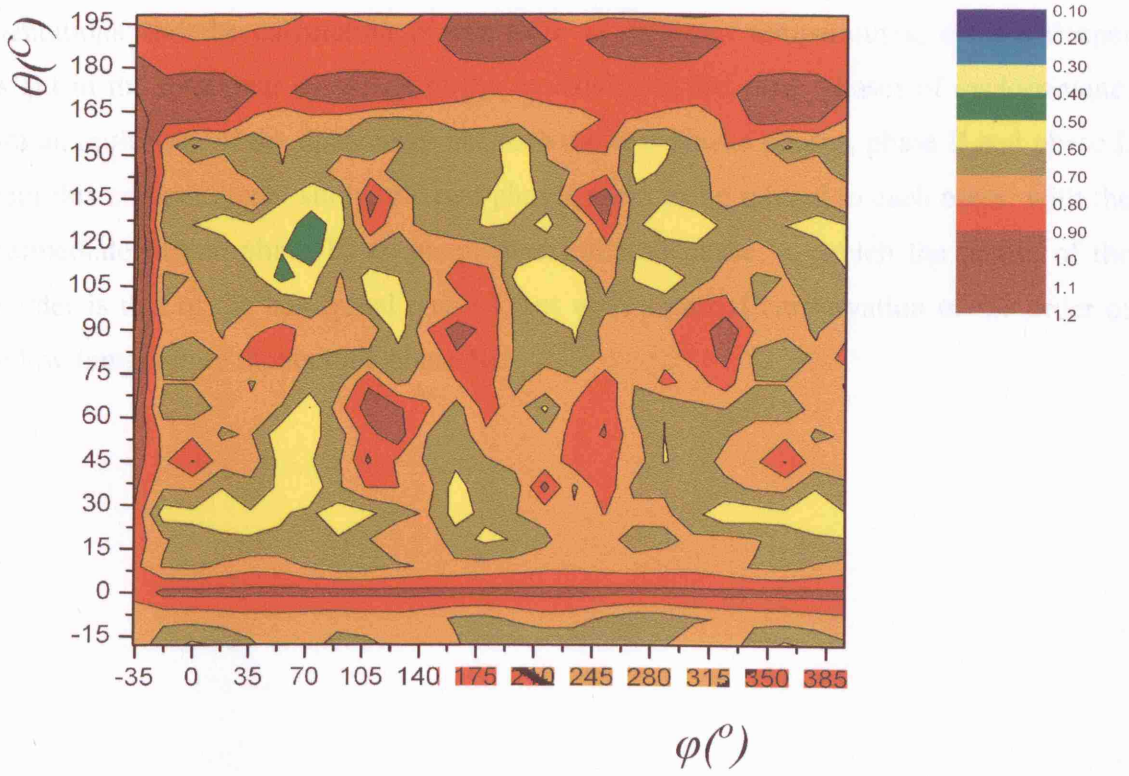


Fig. 7.65: Contour map of the ODF of the *directional* vector v_1 of a molecule in the super-heated hexagonal plastic phase I of cyclopentane at 230K.

The high density of orientational probability is associated with a large variety of other orientational states (θ; φ), indicated by the yellow and dark yellow colour on the contour map (see the legend in figure 7.65). This feature corresponds to a high probability for wide re-orientational motion of the molecules. This ODF reflects the behaviour of a system with virtually free rotating molecules, with an increase in the orientational disorder, which becomes close to that typical of a liquid.

7.4 Conclusions

The MD simulation of the solid phase of cyclopentane, within a rigid model, gives a good reproduction of the experimental results for the low-temperature solid phase III and for the hexagonal plastic phase I. It also simulates correctly the solid-plastic phase transitions, phase III - phase II - phase I, with transition temperatures which are reasonably close to experiment. Because of the lack of experimental data for the intermediate phase II, it was not possible to estimate well the quality of the MD

simulation of the intermediate phase II. For the latter, the calculated XRD powder patterns differ significantly from experiment. The statistical analysis of the molecular orientations and the calculation of the ODF at different temperatures, offer a deeper insight in the rotational dynamics of the molecules in the three phases of cyclopentane, with an indication of the differences between the two plastic phases, phase II and phase I. From this computational study, the two phases seem to be related to each other, with the intermediate plastic phase II, an incipient disordered phase, in which the nature of the disorder is that of the hexagonal phase I, but with a partial conservation of the order of the low-temperature monoclinic phase III.

7.5 References

(1) Boese, R. *unpublished results.*

(2) Torrisi, A.; Leech, C. K.; Shankland, K.; Ibbersen, W. D. R.; Benet-Buchholz, J.; Boese, R.; Leslie, M.; Catlow, C. R. A.; Price, S. L. *(in preparation)* **2007.**

(3) Murugan, N. A. *J. Phys. Chem. B* **2005**, *109*, 23955.

(4) Windsor, C. G.; Saunderson, D. H.; Sherwood, J. N.; Taylor, D.; Pawley, G. S. *J.Phys. C:Solid State Phys.* **1978**, *11*, 1741.

(5) Cardini, G. *Chem. Phys.* **1995**, *193*, 101.

(6) Pawley, G. S.; Dove, M. T. *J.Phys. C: Solid State Phys.* **1984**, *17*, 6581.

(7) Pawley, G. S.; Dove, M. T. *J.Phys. C: Solid State Phys.* **1983**, *16*, 5969.

(8) Yashonath, S.; Rao, C. N. R. *J. Phys. Chem.* **1986**, *90*, 2552.

(9) Cardini, G.; Ricci, M.; Righini, R.; Califano, S. *Chem. Phys.* **1994**, *189*, 17.

(10) Schettino, V.; Marzocchi, M. P.; Califano, S. *J. Chem . Phys.* **1969**, *51*, 5264.

(11) Mack, J. W.; Torchia, D. A. *J. Phys. Chem.* **1991**, *95*, 4207.

(12) Smith, W.; Forester, T. R. *J.Mol.Graphics* **1996**, *14*, 136.

(13) Haile, J. M. *Molecular Dynamics Simulation:Elementary Methods*; Wiley & Sons: United States, 1992.

(14) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquid*; Oxford Science: Oxford, 1987.

(15) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.

(16) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J.Am.Chem.Soc.* **1995**, *117*, 5179.

(17) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(18) Gale, J. D.; Rohl, A. *Mol. Simul.* **2003**, *29*, 291.

(19) Berendsen, H. J. C.; Postma, J. P. M.; VanGusteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(20) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.

(21) Gonzalez-Melchor, M.; Orea, P.; Lopez-Lemus, J.; Bresme, F.; Alejandre, J. *J. Chem. Phys.* **2005**, *122*.

(22) Wiberley, S. E.; Bunce, S. C.; Bauer, W. H. *Anal. Chem.* **1960**, *32*, 217.

(23) Plyler, E. K.; Acquista, N. *J.Res. Natl. Bur. Std.* **1949**, *43*.

(24) Post, B.; Schwartz, R. S.; Fankuchen, I. *J. Am. Chem. Soc.* **1951**, *73*, 5113.

(25) Aston, J. G.; Fink, H. L.; Schumann, S. C. *J.Am.Chem.Soc.* **1943**, *65*, 341.

(26) Velardez, G. F.; Alavi, S.; Thompson, D. L. *J. Chem. Phys.* **2003**, *119*, 6698.

(27) Cahn, R. W. *Nature* **1986**, *323*, 668.

(28) Maddox, J. *Nature* **1987**, *330*, 599.

(29) De Koning, M.; Antonelli, A.; Yip, S. *J. Chem. Phys.* **2001**, *115*, 11025.

(30) Lu, K.; Li, Y. *Phys. Rev. Lett.* **1998**, *80*, 4474.

(31) Jin, Z. H.; Gumbsch, P. *Phys. Rev. Lett.* **2001**, *87*, 055703.

(32) Sprik, M.; Rothlisberger, U.; Klein, M. L. *J. Phys. Chem. B* **1997**, *101*, 2745.

(33) Depondt, P.; Breymann, W. *Mol.Phys.* **1996**, *87*, 1015.

(34) Staveley, L. A. *Annu. Rev. Phys. Chem.* **1962**, *13*, 351.

(35) Kawai, N. T.; Gilson, D. F. R.; Butler, I. S. *J. Phys. Chem.* **1992**, *96*, 8556.

(36) Kawai, N. T.; Gilson, D. F. R.; Butler, I. S. *J. Phys. Chem.* **1990**, *94*, 5729.
(37) Kinoshita, T.; Mashimo, T.; Kawamura, K. *J.Phys.: Condens. Matter* **2005**, *17*, 1027.
(38) Mnyukh, Y. *Fundamentals of Solid State Phase Transitions, Ferromagnetism and Ferroelectricty*, 2001.
(39) Pardo, L. C.; Lunkenheimer, P.; Loidl, A. *J. Chem. Phys.* **2006**, *124*.
(40) Luty, T.; Rohleder, K.; Lefebvre, J.; Descamps, M. *Phys. Rev. B* **2000**, *62*, 8835.
(41) Brunelli, M.; Wright, J. P.; Vaughan, G. B. M.; Mora, A. J.; Fitch, A. N. *Angew. Chem. Int. Edit.* **2003**, *42*, 2029.
(42) Shankland, K.; David, W. I. F.; Silvia, D. S. *J.Mater.Chem.* **1997**, *7*, 568.
(43) Giacovazzo, C.; Monaco, H. L.; Artioli, G.; Viterbo, D.; Ferraris, G.; Gilli, G.; Zanotti, G.; Catti, M. *Fundamentals of Crystallography*; Oxford University Press, 2002.
(44) Kuchta, B.; Descamps, M.; Affouard, F. *J. Chem. Phys.* **1998**, *109*, 6753.
(45) Suryanarayana, C.; Norton, M. G. *X-Ray Diffraction:a practical Approach*; Plenum, Press: New York and London, 1998.
(46) Warren, B. E. *X-Ray Diffraction*; Addison-Wesley: Reading, 1969.
(47) Pawley, G. S.; Refson, K. *Acta Cryst. Sec. A* **1987**, *43*, 727.
(48) Smith, W.; Clarke, J. H. R. *J.Chem.Phys.* **1989**, *90*, 6610.
(49) Meyer, M.; Ciccotti, G. *Mol. Phys.* **1985**, *56*, 1235.
(50) Edmonds, A. R. *Angular Momentum in Quantum Mechanics*Princeton, 1957.
(51) Abramovitz, M.; Stegun, I. A. *Laplace's Equation: Handbook of Mathematical functions with formulas, graphs and mathematical tables*New York: Dover, 1972.
(52) Krantz, S. G. *"The Laplace Equation": Handbook of Complex Variables*Boston, 1999.

Chapter 8

SUMMARY AND CONCLUSIONS

8.1 Summary and Comments of the overall study

The study presented in this thesis focus on the crystal structure prediction (CSP) of organic simple molecules and the simulation of the phase transition among three different polymorphs of cyclopentane, with the use of computational techniques.

The study shows the high complexity of CSP of organic compounds, even in the case of small, simple molecules, such as imidazole, azetidine and cyclopentane. The success in the prediction of the experimental crystal structure of the first compound (chapter 4) confirms the huge progress achieved in the computational CSP of small organic rigid molecules, containing polar groups, for which the electrostatic contribution dominates the intermolecular interactions in the formation of the different polymorphs[1,2]. Our study has confirmed that the use of the distributed multipole analysis[3,4] in describing the intermolecular electrostatic contribution is very important for a successful CSP for small rigid polar molecules, such as imidazole.

Because of the lack of a powerful, accurate and efficient computational method of CSP, based on pure *ab initio* calculations, our method uses intermolecular potentials for the simulation of the intermolecular forces among the molecules in the crystal structure, and it has shown the big importance of the latter for a successful CSP[5-7]. The inadequate intermolecular potential turned out to be one of most determining factors in the failure observed in CSP of azetidine (chapter 5). Nevertheless, the latter system has shown such unexpected high complexity that even the availability of a more accurate potential would probably have not given a successful CSP[2]. Indeed, the presence of a metastable experimental structure, whose nature was discovered by computational analysis, confirmed that a computational method, purely based on static lattice energy minimisation, is not sufficient for a fully successful prediction and a more complex model needs to be developed, which takes into account other important factors such as kinetics and interactions with solvents[5,8].

The study of the crystal structure of cyclopentane (chapter 6) shows the high complexity, which can emerge in the CSP of an apparently simple system, especially when the crystal structures are determined by very weak intermolecular interactions, such as dispersive

forces. Indeed, the presence of a huge amount of local minima in a very small range of lattice energy (~5kJ mol^{-1}) makes the CSP impossible. The latter is a characteristic common among organic compounds in the organic solid state and is one of the major problems in CSP[9]. Nevertheless the method gave interesting structural and thermodynamic information, which can be used in predicting the existence of a plastic phase for an organic compound. This aspect supports once more the important, complementary role of computer simulation in interacting with experimental investigations[7,10,11].

The study of the order-disordered phase transition in cyclopentane solid state, using Molecular Dynamics (MD) simulations (chapter 7), has shown a good quality in the reproduction of the main solid-plastic phase transformation and a good description of the nature of the system in the different phases, which confirms the good capability of this method in simulating specific types of phase transitions, which are not characterised by the presence of high energy barriers between two polymorphs[12,13]. The result is even more outstanding, considering that we used a very simple rigid model for a high flexible molecule, such as cyclopentane. Moreover, through the use of specific additional features, such as X-ray powder patterns and Orientational Distribution Function (ODF) calculation, MD simulations gave us a deeper insight in the nature and the rotational dynamics of the different solid phases of cyclopentane, showing again a useful interaction with experiments.

Finally, the MD study on imidazole (chapter 4, section 4.4), has shown that the MD method can be also useful in CSP, especially in the attempt to establish which of the predicted low energy structures are effectively possible polymorphs. Indeed, MD simulation allows us to take into account the thermal effects, which are not present in a pure static lattice energy minimisation method of CSP, and which play an important role in a such complex field as the polymorphism in the organic solid state[14].

8.2 References

(1) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2004**, *4*, 1327.

(2) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. *Acta Cryst. Sec. B* **2005**, *61*, 511.

(3) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(4) Stone, A. J.; Alderton, M. *Mol.Phys.* **1985**, *56*, 1047.

(5) Gavezzotti, A. *CrystEngComm* **2002**, *4*, 343.

(6) Gavezzotti, A. *Molecular Aggregation*; Oxford Science Publications, 2007.

(7) Price, S. L. *Adv. Drug Deliv. Rev.* **2004**, *56*, 301.

(8) Gavezzotti, A. *Modelling Simul. Mater. Sci. Eng.* **2002**, *10*, R1.

(9) Dunitz, J. D. *Chem. Comm.* **2003**, 545.

(10) Hulme, A. T.; Price, S. L.; Tocher, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 1116.

(11) Lancaster, R. W.; Karamertzanis, P. G.; Hulme, A. T.; Tocher, D. A.; Covey, D. F.; Price, S. L. *Chem. Comm.* **2006**, 4921.

(12) Tuble, C. S.; Anwar, J.; Gale, J. D. *J.Am.Chem.Soc.* **2004**, *126*, 396.

(13) Yashonath, S.; Rao, C. N. R. *J. Phys. Chem.* **1986**, *90*, 2552.

(14) Gavezzotti, A. *J.Am.Chem. Soc.* **2000**, *122*, 10724.

APPENDIXES

Appendix A: Ewald Summation

The Ewald Summation is a method to treat the electrostatic contribution in a periodic system of interacting ions or molecules. The infinite periodic system is divided into different boxes, each of which contains N interacting ions. As such we can express the coulombic electrostatic contribution to the potential as:

$$U_{electr} = \frac{1}{2}\sum_{\boldsymbol{n}=0}^{\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{|\boldsymbol{r}_{ij}+\boldsymbol{n}|} \quad , \qquad [A.1]$$

where q_i, q_j are the charges of the ions, $\boldsymbol{r}_{ij}$ the relative distance between a pair of ions and $\boldsymbol{n}$ is the vector, which defines the position of the boxes, in which the solid is divided. If L is the length of the box, then $\boldsymbol{n}$ = (n_xL, n_yL, n_zL), where n_x, n_y, n_z, are integers.

From equation [A.1] we see that the electrostatic energy is calculated as an infinite series with half of the terms positive, and half of the terms negative. This series is conditionally convergent and depends on the order the terms are summed. To solve this problem Ewald[1] developed a method, which essentially consists of splitting the series into two rapidly convergent series, one in real space and one in reciprocal space.

$$U_{electrost} = U_{real} + U_{reciprocal} \quad . \qquad [A.2]$$

By surrounding each point charge by a Gaussian charge distribution of equal and opposite sign, centred on the point charge, it is possible to screen the point charges, making the interactions between the latter lower. The gaussian function is of the kind:

$$g_i(r) = \frac{q_i \alpha^3}{\pi^{\frac{3}{2}}} e^{-\alpha^2 r^2} \quad . \qquad [A.3]$$

In this way we obtain a first rapidly convergent series, given by the sum of the point charges plus the screening gaussians functions. This series is summed in real space:

$$U_{real} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\boldsymbol{n}=0}^{\infty}\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{|\boldsymbol{r}_{ij}+\boldsymbol{n}|} erfc(\alpha|\boldsymbol{r}_{ij}+\boldsymbol{n}|), \qquad [A.4]$$

where *erfc(x)* is the complementary error function, with the general expression:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt \quad , \qquad [A.5]$$

which is illustrated in figure A.1:

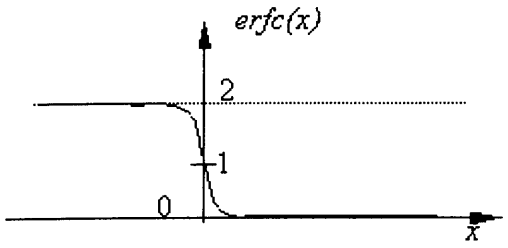


Fig. A.1: plot of *erfc(x)*

The latter function is complementary to the *erf(x)* function which is reported in equation [A.6], and illustrated in figure A.2:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} dt \quad . \qquad [A.6]$$

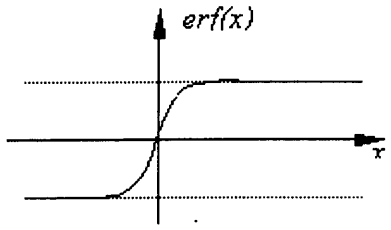


Fig. A.2: plot of the function *erf(x)*.

Between these two functions there is the relation:

$$erf(x) + erfc(x) = 1 \quad . \qquad [A.7]$$

In order to maintain charge neutrality it is necessary to neutralise the charge distribution introduced with the Gaussian functions, which is done by adding other Gaussians with equal and opposite sign to the previous ones. This second charge distribution is easier to calculate in reciprocal space and is also rapidly convergent:

$$U_{reciprocal} = \frac{1}{2} \sum_{\boldsymbol{k} \neq 0}^{\infty} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{4\pi^2}{k^2} e^{-\frac{k^2}{4\alpha^2}} \cos(\boldsymbol{k} \cdot \boldsymbol{r}_{ij}) \quad , \qquad [A.8]$$

where $\boldsymbol{k}$ is the vector of the reciprocal lattice:

$$\boldsymbol{k} = \frac{2\pi \boldsymbol{n}}{L^2} \quad . \qquad [A.9]$$

In the sum in real space an interaction of the Gaussian function with itself is included. It is thus necessary to add a compensating factor in the $U_{electrost}$ series. This term is known as the self-energy contribution:

$$U_{self-energy} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{N} \frac{q_i^{\,2}}{4\pi\varepsilon_0} \quad . \qquad [A.10]$$

In practice the Ewald method is controlled by three parameters: the α parameter in the gaussian function (10^{-6} in our calculations of cyclopentane), a cut-off in the real space summation for the distance $\boldsymbol{r}$ ($\boldsymbol{r}_{cut\text{-}off}$) and a maximum value for the reciprocal vector $\boldsymbol{k}$ in the sum in the reciprocal space ($\boldsymbol{k}_{max}$).

Figure A.3 shows a schematic representation of the method for a 1D lattice.

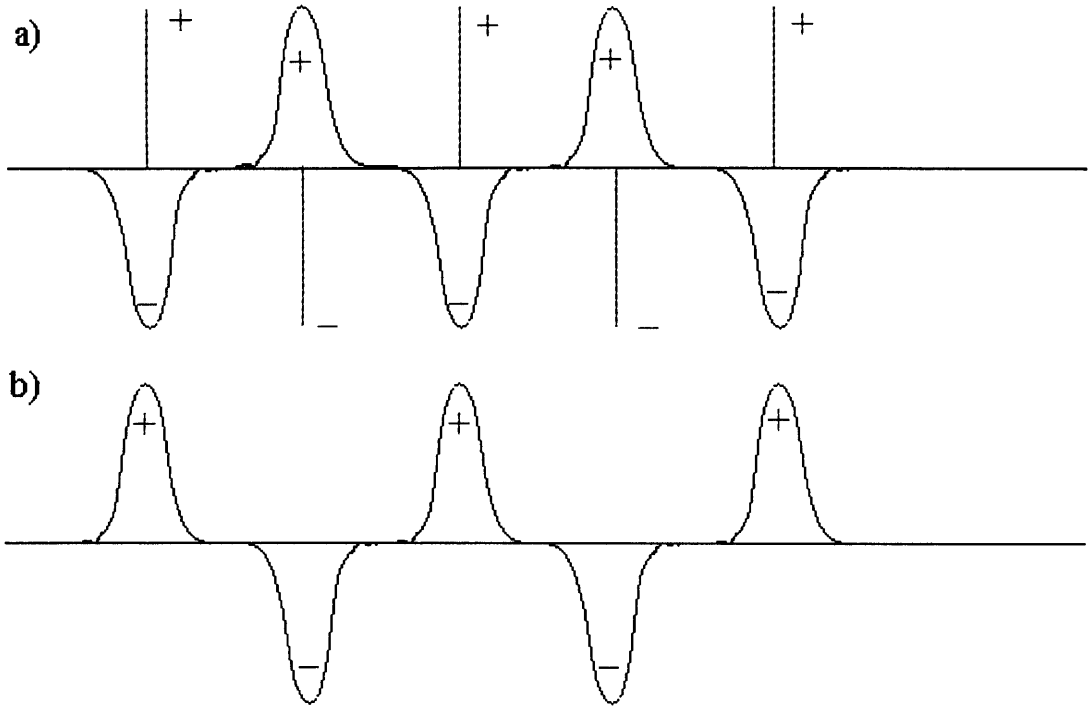


Fig. A.3: Ewald summation method. **a)** Gaussian functions centred on point charges, summed in the real space. **b)** compensating Gaussian functions summed in the reciprocal space.

The Ewald parameter η is used to define the cut-off for the sum in the direct space ($r_{cut\text{-}off}$) and the sum in the reciprocal space (k_{max}). The Ewald parameter is defined in terms of the number of atoms N and the unit cell volume V, by the equation:

$$\eta = \left(\frac{\pi^3 N}{V^2} \right)^{\frac{1}{6}} \qquad [A.11]$$

The default accuracy in the Ewald sum is 10^{-6} and the parameter f is defined as:

$$f = \sqrt{ln\left(\frac{1}{10^{-6}} \right)} \qquad [A.12]$$

The cut-off distance in the direct space and in the reciprocal space are expressed as function of the two parameters f and η by the following expressions:

$$r_{cut-off} = \frac{f}{\eta} \qquad [A.13]$$

$$k_{\max} = \frac{f \cdot \eta}{\pi} \qquad [A.14]$$

Hence, the cut-off values in the Ewald sum depend on the individual, different structures.

Appendix B:

Calculation of the Elastic Constants and Lattice Phonons

In this section we briefly illustrate the theoretical principles used in calculating the lattice phonon frequencies and the elastic constants of a molecular crystal.

In the rigid body approximation the equation describing the lattice frequencies is[2]:

$$\omega^2(k) \cdot w_\tau^M(k) = \sum_{\tau',N} D_{\tau\tau'}^{MN}(k) w_{\tau'}^N(k) , \qquad [B.1]$$

where w_τ^N is the mass weight displacement of the molecule N along τ direction, k is the wave-vector, $\omega(k)$ are the phonon frequencies and $D_{\tau\tau'}^{MN}(k)$ is the dynamical matrix which has the following expression in function of the instantaneous potential energy (u):

$$D_{\tau\tau'}^{MN}(k) = \frac{1}{\sqrt{(M_{\tau,M} \cdot M_{\tau',N})}} \cdot \sum_{l'} \frac{\partial^2 u^{0l'}}{\partial w_\tau \partial w_{\tau'}} e^{ik \cdot r^{l'}} , \qquad [B.2]$$

where 0 represent the reference unit cell of the crystal and *l'* another generic unit cell. The second derivatives of the potential energy from the Hessian matrix in equation [B.2] are calculated numerically and from them is calculated the dynamical matrix $D_{\tau\tau'}{}^{MN}(\boldsymbol{k})$ for $\boldsymbol{k} = \boldsymbol{0}$. The eigenvalues equation to solve is:

$$\left| D_{\tau\tau'}{}^{MN}(\boldsymbol{k}) - \omega^2(\boldsymbol{k})\delta_{\tau\tau'}\delta_{MN} \right| = 0 \quad , \qquad \text{[B.3]}$$

which is done numerically using the program DMAREL[3].

From the phonon frequencies ω, we can give an approximate estimate of the vibrational entropic contribution and hence to the free energy of the system. The Helmholtz free energy *A* has in fact the expression from the statistical mechanics[4]:

$$A = -k_B T \ ln Z = U_o + k_B T \sum_{k,\nu=1}^{3nN} ln\left(\frac{1}{1 - e^{-\frac{\hbar\omega(k,\nu)}{k_B T}}} \right) \quad , \qquad \text{[B.4]}$$

where *Z* represents the partition function of the system in the canonical ensemble, *n* is the number of atoms and *N* the number of unit cells in the crystal and ν the generic vibrational normal mode in the lattice. U_o represents the internal energy of the system, which is given by the lattice energy and the zero point energy as expressed in equation [B.5]:

$$U_o = U_{lattice} + \sum_{\boldsymbol{k},\nu=1}^{3nN} \frac{1}{2}\hbar\omega(\boldsymbol{k},\nu) \quad . \qquad \text{[B.5]}$$

We can see from equation [B.4] and [B.5] that phonon frequencies are necessary to calculate both the zero point energy and the free energy of a crystal structure.

We can also calculate the elastic constants of a crystal lattice. These constants, assuming an isothermal and reversible deformation process, are defined as[5]:

$$C_{ij} = \frac{\partial}{\partial \varepsilon_j}\left(\frac{\partial u}{\partial \varepsilon_i} \right) \quad , \qquad \text{[B.6]}$$

where ε_i *and* ε_j are the components of the *strain tensor* and *u* is the free energy per unit volume. C_{ij} are the components of a fourth-rank tensor, having the property of symmetry so that:

$$C_{ij} = C_{ji} \quad . \qquad [B.7]$$

Hence the elastic constants are obtained as second derivatives of the potential surface energy at the temperature of 0K. The increase of energy density produced by a homogeneous strain ε_i is called strain energy and has the following expression[5]:

$$\Delta u = \frac{1}{2}\sum_{ij} C_{ij}\varepsilon_i\varepsilon_j \quad . \qquad [B.8]$$

It takes into account the internal and external strain contributions to Δu.

Appendix C: Calculation of the Volume Growth Rate

- ***The BFDH model.*** The BFDH model, taking its name from Bravais-Freidel-Donnay-Harker[6-8], defines a criteria to predict crystal morphology. It is essentially based on considering the growth of the different faces of a crystal. The lowest growth rates of the faces occur for the faces with largest interplanar space d_{hkl}. The faces having the lowest energy are the most important from a morphological point of view. Hence the growth rate in the [*hkl*] plane, R_{hkl} , depends on the inverse of the lattice interplanar distance d_{hkl}:

$$R_{hkl} \propto \frac{1}{d_{hkl}} \quad . \qquad [C.1]$$

- ***Growth Morphology (AE Model).*** This model is based on the Hartman-Perdok[9] theory of morphology, which considers the role, played by the intermolecular forces in the crystallisation process. The crystal faces are classified in three groups, depending on the number of periodic bond chains (i.e. hydrogen bonds) in the growth slice:

  F (flat) face: with two periodic bond chains per slice.

  S (stepped) face: one periodic bond chain per slice.

  K (kinked) face: no periodic bond change per slice.

  Only the F faces can grow with a layer mechanism, whilst S and K faces have a continuous mode of growth. Hence the growth of F face will be the slowest and thus it will be the dominant face from a morphologic point of view.

A key factor of the theory is the attachment energy, which is the energy per molecule released when another slice is attached to the face. It is directly proportional to the growth rate R and hence inversely proportional to its morphological importance. Calling $E_i(hkl)$ the interaction energy between the slice of thickness d_{hkl}, and the i^{th} underlying slice, the attachment energy can be defined as:

$$E_{attach.} = \sum_{i=1}^{\infty} E_i(hkl) \ . \qquad [C.2]$$

For a stable surface the attachment energy must be exothermic. It can be used to represent the *kinetics in vacuum* of the crystal. In fact this relationship is valid:

$$E_{crystal} = E_{slice}(hkl) + E_{attach.}(hkl) \ , \qquad [C.3]$$

where $E_{slice}(hkl)$ is the energy between all atoms in that slice. Since the energy of the crystal ($E_{crystal}$) is constant for all faces, a large slice energy ($E_{slice}(hkl)$) implies a low attachment energy ($E_{attach}(hkl)$), which implies a low value of the growth rate R_{hkl} for the face. Hence this face will be morphologically important.

- ***Equilibrium Morphology (SE Model).*** This model calculates the morphology of a crystal using the surface energy E_{surf}, which is the energy, required to cleave a surface of ions from a bulk material as shown in the figure C.1 for a 2D crystal[10].

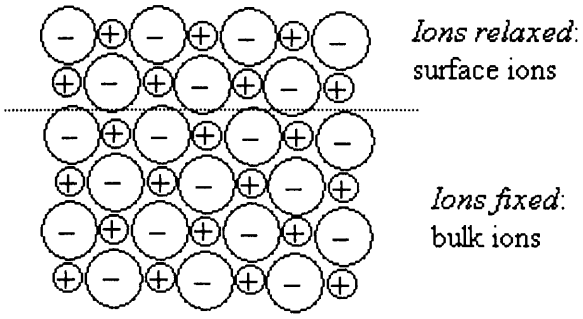


Fig. C.1: 2D ion crystal with the separate regions: surface ions and bulk ions.

The surface energy is hence given by the difference between the energy of the surface ions and the energy of the ions in the bulk per unit surface area. It is defined as:

$$E_{surf.}(hkl) = \frac{[E_{total}(hkl) - E_{boundary}(hkl) - nE_{crystal}]}{A(hkl)} \quad , \qquad [C.4]$$

where E_{total} is the energy of the entire system, n is the number of unit cells on the surface of the crystal, $A(hkl)$ is the area of the face and $E_{boundary}$ is the difference between the energy of the surface atoms and the energy of the bulk atoms for that face. This surface energy has to be a minimum for a crystal of given volume, in equilibrium with the surrounding crystals. It gives a measure of the thermodynamic stability of a surface and it is known as the *equilibrium morphology*. The lower is the surface energy of a face, the lower its growth rate will be and the more important this face will be from a morphological point of view.

If either the AE model or the SE model is used to calculate the crystal morphology, then effects due to the surface relaxation are often ignored. It was noted that the SE model often overestimates the importance of the higher index faces of the crystal in calculating the morphology, compared to the AE model[10].

- ***The Volume Growth Rate.*** The typical method to calculate the growth rate of a crystal is based on the calculation of the attachment energy E_{attach} or the surface energy E_{surf} of a set of faces, properly chosen. Consequently it is possible to draw the morphology for each face using the so called Wulff plot[10]. In this plot (an example of which is given in figure C.2) the ratio of surface normal distances of all planes from the centre of the crystal can be determined.

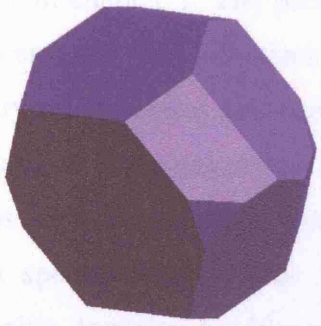


Fig. C.2: example of Wulff-plot for an imidazole crystal structure.

The volume growth rate is obtained by numerical integration to find the volume within the Wulff plot shape. Calling D_{max} the maximum distance between the centre of the crystal and the corners of the Wulff shape and d_{min} the minimum distance between the centre of the crystal and the faces, we can define an *aspect ratio r* as:

$$r = \frac{D_{\max}}{d_{\min}} \quad . \qquad [C.5]$$

The larger E_{attach}., the larger the volume and the faster the crystal grows. By contrast the smaller E_{attach}, the smaller the volume and the slower the crystal grows.

Although the method is approximate, because the modelling takes place in vacuo and neglects the influence of the solvent, the calculation of relative volume growth rates can be very helpful in the attempts of predicting polymorphism. Indeed such calculations can suggest which thermodynamically feasible crystal structures could have a kinetic advantage in crystal growth[11]. The method is based on the use of interatomic and intermolecular potentials, but the results are often not very sensitive to the intermolecular potential used in the prediction of the morphology of a crystal[12]. Different programs like MARVIN[13] and HABIT95[14] are used to calculate the morphology of a crystal using the AE model.

Appendix D: FIT parameters of the interatomic potential

We report in table D.1 the semi-empirical FIT parameters of the repulsive-dispersion contribution of the *6-exp* intermolecular potential[15], used in all the studies performed in this work and discussed previously in section 1.5.2. This potential has been already discussed in chapter 1,section 1.5.2. The repulsive-dispersive contribution is a *6-exp* form, reported in equation 1.37 in chapter 1. The parameters *A*, *B*, *C* are required to describe the interaction between each pair of atoms. They are usually derived by empirical fittings to a set of crystal structures and heats of sublimations for different compounds, but also by fitting to *ab initio* calculated data. We report in the table the parameters for the interactions between the same species of atoms. The parameters interactions between different species (i.e. C-N, or N-H) are determined from the parameters in the table, by using apposite combination rules, reported in equation [1.38] in section 1.5.2.

Potential	*Atom Pair*	*A* ($kJmol^{-1}$)	*B* ($Å^{-1}$)	*C* ($Å^6 kJmol^{-1}$)
FIT	C···C	369743	3.60	2439.8
	N···N	254529	3.78	1378.4
	H_C···H_C	11971	3.74	136.4
	H_N···H_N	5030	4.66	21.5

Table D.1: Values of the FIT parameters[15] of the Buckingham *(exp-6)* intermolecular potential used in the computational studies presented in this work. H_C indicates the *non*-polar hydrogen atoms bonded to carbons in a specific molecule, while H_N indicates the polar hydrogens, bonded to nitrogen in a specific molecule.

Appendix E: Computational Resources Information.

The calculations relating to the crystal structure prediction of imidazole, azetidine and cyclopentane compounds were performed on a cluster 64-bit Xeon linux machine with 52 nodes, each of them having 2 Intel processors.

The MD simulations have been performed on HPCx, which is an IBM POWER5 machine, with 2560 processors, delivering 15.36 TeraFlop/s peak and with 5.12 TByte of memory and 72 TByte of disk. The operative system is IBM's AIX.

References

(1) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.

(2) Born, M. H. K. *Dynamical Theory of Crystal Lattice*; Oxford University Press: New York, 1954.

(3) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comput. Chem.* **1995**, *16*, 628.

(4) Stowe, K. S. *An introduction to thermodynamics and statistical mechanics 2nd Ed.*; Cambridge University Press, 2007.

(5) Nye, J. F. *Physical Properties of Crystals*, 1985 ed.; Oxford University Press, 1985.

(6) Bravais, A. *Etudes crystallographiques*; Academie des Science: Paris, 1913.

(7) Friedel, G. *Bull. Soc. Fr. Miner.* **1907**, *30*, 326.

(8) Donnay, J. D. H.; Harker, D. *Am. Miner.* **1937**, *22*, 446.

(9) Hartmann, P.; Perdok, W. G. *Acta Cryst.* **1955**, *8*, 49.

(10) Coombes, D. S.; Catlow, C. R. A.; Gale, J. D.; Hardy, M. J.; Saunders, M. R. *J. Pharm. Sci.* **2002**, *91*, 1652.

(11) Coombes, D. S.; Catlow, C. R. A.; Gale, J. D.; Rohl, A. L.; Price, S. L. *Cryst. Growth Des.* **2005**, *5*, 879.

(12) Brunsteiner, M.; Price, S. L. *Cryst. Growth Des.* **2001**, *1*, 447.

(13) Gay, D. H.; Rohl, A. L. *J. Chem. Soc.-Faraday Trans.* **1995**, *91*, 925.

(14) Clydesdale, G.; Roberts, K. J.; Docherty, R. *J. Cryst. Growth* **1996**, *166*, 78.

(15) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352.